AN ASSESSMENT OF THE CONSTRUCT VALIDITY OF INFANT TEMPERAMENT RATINGS USING MATERNAL DIARIES

By

DONNA S. KITCH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

This work is dedicated to my two children, Lacey and Sam, who have demonstrated to me repeatedly that there really is such a thing as temperament.

ACKNOWLEDGEMENTS

I would like to thank the all of the mothers who participated in this study for inviting me into their homes and taking the time to record their infant's behavior over four days. Without their interest and cooperation this study would not have been possible. I would also like to thank the research assistants who helped me with interviewing and coding: Heidi Williams and Andrea Berger deserve special acknowledgement for their dedication and consistent effort.

I would also like to thank the members of my committee for their patience and support throughout this project. Very special thanks are extended to the Chairman of my committee, Dr. Franz Epting. From my first year in graduate school onward, he has guided my professional development and has served as my mentor in construing and refining my own philosophical orientation. He has consistently stood by me through many changes and has advised me wisely while trusting my own judgement and intuition. I would also like to thank Dr. Margaret Wilson, who joined the committee on very short notice and offered a generous amount of assistance and advice during the last few months of work on this project. Thanks are also extended to Dr. Yvonne Brackbill, who has

consistently supported and guided me throughout my years in graduate school. She has taught me a great deal and for that I am grateful.

Finally, I would like to thank two people who have been invaluable to me throughout this effort. Helen Cunningham has been not only a respected colleague but a true friend, who was always there when I needed her. And very special thanks are extended to my husband Gary, who has demonstrated incredible patience throughout my years in graduate school and whose persistent faith in me has given me the courage to succeed.

TABLE OF CONTENTS

LIST OF TABLES

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

AN ASSESSMENT OF THE CONSTRUCT VALIDITY OF INFANT TEMPERAMENT RATINGS USING MATERNAL DIARIES

By

Donna S. Kitch

August 1993

Chairperson: Dr. Franz Epting
Major Department: Psychology

Research demonstrating the correlation between maternally rated infant temperament and maternal characteristics has cast doubt on the construct validity of questionnaires using maternal ratings of infant temperament. This criticism has been most specifically focused on the Revised Infant Temperament Questionnaire (RITQ). RITQ ratings of infants on the five scales of Approach, Adaptability, Mood, Rhythmicity and Intensity obtained from 45 mothers were compared with their Diary reports of the infant's behavior over four days. The Infant Characteristics Questionnaire (ICQ) and observer ratings were also used for comparison. Only two of the five RITQ dimension scores correlated at a significant level with corresponding Diary dimension scores, though four of the Diary scales correlated with corresponding ICQ factors. Observer ratings did not correlate with either RITQ or ICQ scales,

although correlations between the observer scores and the corresponding Diary scores approached significance. RITQ dimension scores did correlate at a moderate level with corresponding ICQ factor scores. Thus the construct validity of the RITQ was partially supported. Results were attributed to both the psychometric weaknesses of the RITQ and to the possibility that RITQ ratings reflect maternal perceptions of infant behavior, which are likely to be influenced by maternal characteristics and attitudes.

INTRODUCTION TO THE PROBLEM

The New York Longitudinal Study (Thomas, Chess, & Birch, 1968) and its conclusions that the temperament of the child plays an important role in his or her adjustment within the family context led researchers to begin to study infant and child temperament as an important variable in the parent-child relationship and in the development of behavior problems in children. As a result of the work of Thomas and Chess, parent questionnaires involving rating scales have been developed to measure infant and child temperament, and have been widely used in researching the relationship of temperament to other variables of interest. Despite the popularity of these measures in research and clinical use, however, the validity of these questionnaires as indicators of an infant's temperament has yet to be firmly established. Perhaps the most salient criticism of maternal rating scales for the assessment of infant temperament has arisen from the finding that these maternal ratings have been demonstrated to correlate as highly with maternal demographic and psychological variables as with observed infant behavior. The

question then arises as to what temperament questionnaires are actually measuring. The current study aims to provide validity data on the Revised Infant Temperament Questionnaire, which is the most popular measure of infant temperament based on the New York Longitudinal Study conceptualization of temperament, using the comparison measure of Diary reports of infant behavior.

REVIEW OF THE LITERATURE

Conceptualization of Temperament

The exact definition of the term "temperament" has been one of the most controversial issues in temperament research. There is no clear consensus regarding the nature of the construct of temperament. Despite this lack of consensus, temperament research continues at an "exponential pace" (Bates, 1986). Strelau and Angleitner (1991) comment that this lack of agreement is typical in psychology; they state that "there is no consensus among psychologists in the understanding of most of the concepts in psychology." (page 3). Lack of consensus on definition does not invalidate research findings, nor negate the clinical or theoretical significance of results; the field of research on intelligence is an example.

Bates (1987) writes that the most generally agreed upon definition of temperament is that "it consists of biologically rooted individual differences in behavior tendencies that are present early in life and are relatively stable across various kinds of situations and over the

course of time" (page 1102). He goes on to say that there is general agreement that temperament is most often manifest in the context of social interaction, and that it is most often applied to the behavioral aspects of emotion, attention and activity. He sees the central definitional issue as disagreement in the extent to which temperament is defined as surface behavior versus underlying organization, that is the extent to which temperament is viewed as encompassing observable behavior versus underlying biochemical, constitutional or genetic factors. (For example, the approach of Thomas and Chess focuses entirely on patterns of observable behavior, and makes no assumptions or conjectures about the underlying causes of that behavior. In contrast, the approach of Buss and Plomin emphasizes the heritability of characteristics as one of the inclusion criteria for temperamental characteristics.) Bates further speculates that it may be helpful to think of the concept of temperament as encompassing three different levels. The first level is the conceptualization of temperament as a pattern in observed behavior. At this level, the focus of study is the individual's observable patterns of behavior. The second level of conceptualization views temperament in terms of factors of neurological individuality. This level emphasizes differences in anatomical and functional patterns of the central nervous system. The third level of conceptualization of temperament encompasses constitutional

factors, most particularly genetic influences, though also including prenatal environmental influences. Much of the variability among approaches to the study of temperament exists in the differential emphasis on one or more of these levels; this emphasis is evident in the varying definitions of temperament and in the variables utilized to study temperament. However, researchers generally agree that temperament involves some combination of biological and environmental factors, and is expressed in observed behavior.

In a recent roundtable discussion on temperament (Goldsmith et al., 1987) eight prominent researchers outlined points of consensus and points of disagreement in the participants' approaches to the definition of temperament. Points of consensus included the following: 1.) that temperamental dimensions reflect behavioral tendencies rather than discrete behavioral acts; 2.) that temperament does have biological underpinnings; 3.) that temperament has a certain amount of continuity over time within the individual; 4.) that the link between temperament and behavior becomes more complex over time, and that it is relatively direct only during infancy; and 5.) that temperament refers to individual differences rather than species-general characteristics. Points of disagreement were that: 1.) each approach sets a different boundary in terms of how much of behavior is temperament; 2.) that

different dimensions of behavior are included in the definitions; 3.) that approaches differ as to how much of personality is subsumed by temperament; and 4.) that disagreement exists over whether or not the concept of "difficultness" should be retained.

Approaches to the Study of Temperament

Of the many different conceptualizations of temperament, the four approaches of Goldsmith, Buss and Plomin, Rothbart, and Thomas and Chess are the most widely used in infant temperament research in the United States. These approaches will be summarized individually here, with particular emphasis on the Thomas and Chess conceptualization, which is the approach of interest in this study.

Goldsmith: An Emotion-Centered Approach

H. Hill Goldsmith, in collaboration with Joseph Campos, has proposed a definition of temperament as "the set of characteristic individual differences in the intensive and temporal response parameters of behavioral expression of affect-related states" (Goldsmith & Campos, 1982 p.189). They see temperament as emotional in nature, and as indexed by the behaviorally expressed aspects of emotion. Emotion is so central to their conceptualization that Goldsmith asserts that they could substitute the term emotionality for temperament in their research (Goldsmith et al., 1987). They define emotions as "feeling states with their associated

central nervous system states which serve both to motivate the individual, and unless blocked from behavioral expression, to communicate socially significant information to others in the environment" (Goldsmith & Campos, 1982, p.177).

According to Goldsmith, temperament and personality are basically the same; their differences lie primarily in the degree to which the influences of social relations and self-concept are salient. Operating under the assumption that infants are less susceptible to the influences of socialization, and their behavior is less strongly moderated by cognitive processes, their study of temperament is confined to the infancy period (Goldsmith, Elliot & Jaco, 1986).

Buss and Plomin: The Criterial, or Genetic Approach

Arnold Buss and Robert Plomin together define temperament as a set of inherited personality traits that appear early in life (Buss & Plomin, 1984). Of central importance to their approach is that temperamental traits be genetic in origin: they exclude personality characteristics that are assumed to be primarily environmentally determined. They also emphasize that temperamental characteristics are reflected in later personality traits, thus excluding dimensions accepted by other researchers that do not seem to have a significant impact on later personality (such as rhythmicity), or traits that are not generally considered to

be personality variables (i.e. aspects of cognition). According to Buss and Plomin, for a trait to be considered a temperamental one it must be heritable, stable, and predictive of adult personality. Their approach utilizes four broad dimensions of temperament: emotionality, activity, sociability, and impulsiveness. The EASI parent rating questionnaire was designed to measure these four dimensions.

Rothbart: the Biosocial Approach

The approach of Mary Rothbart and her colleagues views temperament as biologically based, and as reflecting not simply genetic factors but the complete biological makeup of the individual, which is influenced not only by heredity but by life experience and maturation (Rothbart, 1986). Their approach has been influenced by models of neurophysiological development, and has defined temperament as relatively stable, constitutional differences in reactivity and self-regulation. Reactivity is here defined as "the propensity toward emotional, attentional and motor responses to stimulation" (Rothbart, 1988, p.1241), and is operationalized by observation of behaviors related to motor activity, vocal activity, and emotional expression such as smiling, laughter, fear and frustration. Self-regulation refers to processes functioning to modulate reactivity. Measurable behaviors reflecting self-regulation include those related to attentional regulation, self-soothing, and

approach-avoidance. Rothbart and her colleagues prefer a multi-method approach, utilizing not only the questionnaires they have developed (the IBQ in infancy), but also home and laboratory observation (Rothbart & Goldsmith, 1985).

Rothbart sees temperament as forming the biological base for the developing personality; but notes that temperament is itself developing. Like Goldsmith, she sees infancy as the ideal period in which to study temperament, although her research encompasses all age groups, including adults.

Thomas and Chess: The Behavioral Style Approach

The most widely used definition of temperament is that developed by Thomas and Chess. Thomas and Chess conceptualize temperament as the stylistic component of behavior, the how of behavior, versus the why of behavior, as in motivation. They explain that a group of individuals may all have the same motives for performing a behavior but may differ markedly in the way they perform that behavior. This approach is based on their extensive New York Longitudinal Study, in which they interviewed and tested infants, children and their parents in 141 families extensively over a period of six years (Thomas & Chess, 1977). Unlike the other temperament theories, Thomas and Chess's conceptualization is primarily descriptive. It does not attempt to explain the mechanisms underlying temperamental traits. Thomas and Chess, and most of the

proponents of their model, come from an applied science perspective; they are primarily clinicians (i.e. pediatricians, child psychiatrists, clinical psychologists) rather than basic researchers.

Thomas and Chess view temperament as the child's contribution to the interactive process between parent and child that eventually shapes personality. Their research was revolutionary in that, unlike other developmental and clinical researchers at the time, they viewed the child as an active agent from the moment of birth onward in the developmental process. The concept of "goodness of fit" is central to their approach: they assert that of primary importance in the parent-child relationship is the goodness of fit between the child's temperament and the parent's characteristics and expectations. For example, an infant who is temperamentally very active and intense will be accepted more readily by a parent who is active and values robust activity than by a more quiet, passive parent who prefers less stimulation. Similarly, a parent who has a very busy work schedule and values organization and predictability will adjust more readily to an infant who is rhythmic, or predictable, in sleep, eating, and bowel patterns than to one who is less rhythmic.

The conceptualization of Thomas and Chess divides the structure of temperament into nine dimensions, derived from an inductive content analysis of their interviews with

parents of infants. These dimensions are Activity, Mood, Approach, Adaptability, Rhythmicity, Intensity, Persistence, Distractibility, and Threshold. Activity refers to the motor component of a child's functioning; the extent to which a child is physically active. In scoring this category, information was gathered as to activity level during bathing, feeding, playing, dressing and handling. Information was also gathered about sleep patterns, and the extent to which the child actively reaches, crawls and walks. Mood refers to the "amount of pleasant, joyful and friendly behavior, as contrasted with unpleasant, crying, and unfriendly behavior". (Thomas & Chess, 1977, p.21.) Approach/Withdrawal refers to the nature of the subject's initial response to a new stimulus. (e.g. new food, toy, or new person) Adaptability refers to eventual responses to new or altered situations. The focus here is not the nature of the initial response but with the ease in which that response can be modified in desired directions. Rhythmicity refers to the predictability and/or unpredictability in time of any function. Behaviors judged to be occurring in regular or irregular patterns include sleep, hunger, and elimination. Intensity refers to the energy level of a response, regardless of its quality, or positive or negative direction. Distractibility refers to how effective extraneous environmental stimuli are in interfering with or in altering ongoing behavior. Threshold refers to "the

intensity level of stimulation that is necessary to evoke a discernible response irrespective of the specific form that the response may take or the sensory modality affected" (Thomas & Chess, 1977, p.21). Behaviors utilized in assessing threshold are those concerning reactions to sensory stimuli, environmental objects, and social contacts.

From these nine dimensions, Thomas and Chess derive four temperamentally based typological characterizations of infants: the easy child (seen in 40% of their sample), the difficult child (10% of their sample), the slow to warm up child (15% of the sample), and the intermediate child (the remainder of the sample, that did not fall into one of the three above categories). Easy infants are characterized by high rhythmicity, positive mood, high approach, high adaptability, and low intensity. Difficult infants are characterized by the opposite pattern: low rhythmicity, negative mood, low approach, low adaptability, and high intensity. The slow-to-warm-up child is characterized by low activity, withdrawal, low adaptability, negative mood, and low intensity.

Temperament Concepts, or Dimensions Common Among Approaches

Although different conceptualizations utilize different dimensions of behavior in their definition and study of temperament, there are several concepts, or dimensions, that are common to more than one approach. Bates (1989) presents a list of "specific temperament concepts for which there

exist relatively extensive construct validational findings" (p. 8). He notes that the operational definitions of these concepts within each approach differ, and therefore probably do not correlate highly with one another, but that they are conceptually similar, and in many cases research evidence has shown that they converge. These concepts include negative emotionality, adaptability, reactivity, activity, attention regulation, sociability and positive emotionality, and the concept of difficult temperament.

The nine Thomas and Chess dimensions are represented in these concepts as follows. Negative emotionality is reflected in Thomas and Chess's Mood dimension, though their dimension includes positive mood as well. Adaptability is reflected in the both the dimensions of Approach/Withdrawal and Adaptability. Reactivity is reflected in the Thomas and Chess dimensions of Threshold and Intensity of response, and activity has its own dimension in the NYLS scheme. Attention regulation is contained in the dimensions of Persistence and Distractibility, and sociability and positive emotionality are encompassed in Mood and, to a certain extent in Approach/Withdrawal and Adaptability. The Thomas and Chess conceptualization, as mentioned previously, has its own definition of the concept of difficultness. The concept of difficult temperament will be described in more detail here, as it has particular relevance to the present study.

The concept of "Difficultness"

The temperament "sub-concept" (Bates, 1980, p. 89) or "cluster of related groups of dimensions" (Prior, 1992, p. 263) of temperament most often used in both clinical practice and in research on the development of behavior disorders is the concept of "difficult" temperament (Bates, 1980). As with the more general term "temperament", the concept of difficultness lacks universal agreement as to its specific components. Despite this lack of agreement on definition, however, there has been a good deal of research demonstrating correlations between difficult temperament and current and future behavioral adjustment in children, thus making it a potentially very important concept. If one can predict from early measures of temperament the likelihood of a child developing behavior problems, intervention can begin early and most likely be more effective.

Thomas and Chess's conceptualization of difficulty as being characterized by withdrawal from novel situations, low adaptability, high intensity, negative mood and low rhythmicity has been the most influential and commonly used. Bates (1980) notes, however, that it is not entirely clear how this conceptualization was derived and for what age groups it is relevant. Other researchers have found more empirically derived characterizations of difficultness more useful; both Bates (Bates, Freeland, & Lounsbury, 1979) and the Australian group of researchers (Prior, Oberklaid &

Sanson, 1989) have developed factor-analytically derived measures of difficultness based on the original Thomas and Chess dimensions.

Of central importance in all definitions and conceptualizations of infant difficultness are the concepts of negative emotionality and the implication of management problems for caregivers. Some researchers have argued that because the concept implies at least a partial origin in social interaction, and has negative value connotations, that it should not be used (Rothbart, 1982). Others have argued that the concept of difficultness is indeed a social construction, and not a within-child characteristic, but that it nevertheless should not be abandoned as a temperament construct because of its high external validity and its clinical usefulness (Bates, 1980).

Thomas and Chess assert that temperamental difficultness is not, as Bates (1980) insists, a social construction but a set of innate temperamental characteristics of the child that prove to be difficult for caregivers. Their argument for the retention of the difficultness concept lies primarily in their belief in its clinical usefulness. By explaining to parents that their child's temperament is inherently difficult to deal with, parents are relieved of blame and can focus on strategies for dealing with their child's behavior. They argue that defining difficultness as a social perception infers that

parents are the origin of the problem, and hence clouds the issue and slows clinical progress.

They acknowledge, however, that "difficult temperament may not impose the same stresses, either in kind or intensity, in different cultures with different types of demands and expectations of the young child" (Thomas, Chess & Korn, 1982; p.18). Temperamental traits perceived as difficult in one culture may be perceived as easy in another (Super & Harkness, 1986). Similarly, perceptions of what behavior is difficult may vary within cultures, according to the situational characteristics and personalities of parents. Additionally, behaviors perceived as difficult at one age may not present difficulty at another age: i.e. distractibility may be a positive characteristic of an infant, whereas it may present a great deal of difficulty for a teacher in a kindergarten classroom.

Methods of measuring temperament

Investigators use many different methods to measure temperament in infancy. Infant temperament can be measured directly by the investigator, through observation in the laboratory or in the home; or indirectly via a respondent, through questionnaire or interview. Each strategy has its advantages and disadvantages (see Rothbart & Goldsmith, 1985). Observational techniques may be preferred because of their assumed objectivity and opportunity for laboratory controls. But they are time-consuming, as long intervals of

time and multiple observations are needed to assure that the behavior observed is representative. Additionally, observational measures, unless conducted unobtrusively in the home environment, are subject to the interference of the effects of the unusual laboratory setting. And some dimensions of temperament, such as rhythmicity, are not at all amenable to measurement by observation.

Indirect measures of temperament also have their advantages and disadvantages. Interviews and questionnaires are relatively easy to administer, and a great amount of data can be collected in a short period of time. However, the bias of the reporter must be taken into account. Some researchers have gone so far as to say that questionnaire measures of temperament are really only measuring maternal or other reporter characteristics (Vaughn, Bradley, Joffe, Seifer & Barglow, 1987). Other researchers point out that the important variable <u>is</u> the mother's perception of the infants temperament. This is the temperament she is responding to and interacting with, and whether or not the infant objectively displays these characteristics is irrelevant. Researchers agree, however, that although there is a certain degree of bias in indirect measures of temperament, the information they provide is valuable, and not obtainable by other means. Therefore the measures should be used, but their biases should be carefully explored and taken into account in research conclusions.

Questionnaires with rating scales remain the most popular choice in the measurement of temperament (Goldsmith & Rieser-Danner, 1990).

The Revised Infant Temperament Questionnaire (RITQ)

The first and most popular of the temperament questionnaires based on Thomas and Chess's nine dimensions was Carey's Infant Temperament Questionnaire (ITQ), a 70-item questionnaire designed for use in his practice of pediatrics, as a screening device. Despite its popularity, this instrument was weak psychometrically, and criticized for its low internal consistency and lack of discriminant validity (Campos, Barrett, Lamb, Goldsmith & Stenberg, 1983). In response to these criticisms, and to make the questionnaire suitable for research purposes as well as clinical use, Carey and McDevitt revised the questionnaire in 1978 to improve its psychometric adequacy. Using a standardization sample of 203 infants from private pediatric practices, Carey and McDevitt (1978) revised the original questionnaire in several ways: they added items, doubled the response choices from three to six, added high-low item reversals, and randomized the items as to situational context. From an item pool of 112, items with correlations of less than .30 with their assigned category were deleted. The Revised Infant Temperament Questionnaire (RITQ) was a 95 item questionnaire with improved reliability. Carey and McDevitt (1978) reported internal consistencies for the nine

scales ranging from .53 to .71, with internal consistency for the overall questionnaire being .83. Test-Retest reliabilities ranged from .66 to .81, with a test-retest correlation for the entire questionnaire of .86 (See Table 1). The standardization sample for the RITQ was a relatively small sample (n=203) of white infants from predominantly upper-middle class families; a fact that has led others to question the scale's validity when used with less affluent and educated populations. Carey and McDevitt's method for categorizing infants into diagnostic clusters was based on means and standard deviations of this sample.

Table 1.

RITQ Normative Sample Descriptive Statistics

Scale	Mean	S.D.	Internal Consistency	Test-Retest
Rhythmicity	2.36	.68	.66	.75
Adaptability	2.02	.59	.57	.74
Intensity	3.42	.71	.56	.66
Mood	2.81	.68	.53	.81
Approach/ Withdrawal	2.27	.78	.71	.77

(Carey & McDevitt, 1978)

Carey and his colleagues have developed three other questionnaires paralleling the RITQ that are designed for use at later ages: the Toddler Temperament Questionnaire

(Matheny, Wilson, & Nuss, 1984), the Middle Childhood Questionnaire (Hegvik, McDevitt, & Carey, 1980), and the Behavioral Style Questionnaire (McDevitt & Carey, 1975). The development of these instruments has increased the utility of the RITQ for longitudinal or comparison studies, since the three measures are comparable in design and in content.

The stability of RITQ scores has been examined in several studies, and is generally found to be moderate (Koniak-Griffin & Rummell, 1988; McNeil & Persson-Blennow, 1988; Peters-Martin & Wachs, 1984). The highest levels of stability are usually attained when the retest interval is no longer than six months, although McDevitt and Carey (1981) report significant and moderate correlations between RITQ scores at 5-8 months and TTQ scores at 1-3 years.

The RITQ has been translated into several other languages and used in China (Chen, Yu, Wang & Tong, 1990), Japan (Hara, Mitsuishi & Yamaguchi, 1990), Greece (Kyrios, Prior & Oberklaid, 1989), Germany (Rennen-Allhoff & Reinhard, 1988), Iceland (Tomasdottir, Wilson, White & Agustsdottir, 1991), and Australia (Sanson, Prior, & Oberklaid, 1985). One study comparing infants with parents from different countries found significant differences between the groups on most dimensions of temperament (Prior, Kyrios & Oberklaid, 1986). They reported that, in general, those infants with parents born in Greece and in Middle

Eastern, North American, and some Asian countries were more likely to show characteristics of difficult temperament than infants with parents from North Western Europe and India.

Validity of the RITQ

According to Barnett and MacMann (1990), "validity evidence is subsumed by the body of research that follows a scale" (p.38). A test's validity focuses on what the test measures, and what generalizations can be made from the results. Research results validate a specific use for a test, not the test itself (Nunnally, 1978); therefore there is no such thing as "high" or "low" validity in a general sense. Validity is typically organized into types, or categories. The categories traditionally utilized in psychological research are criterion-related validity and construct validity (AERA, APA & NCME, 1985).

Criterion-related validity. Criterion-related validity refers to the degree to which a test predicts behavior or classification status on an independent criterion (Barnett & MacMann, 1990). The term concurrent validity refers to studies in which the criterion is measured at the time of testing, and the term predictive validity refers to studies in which the criterion is measured in the future.

Studies assessing the RITQ's concurrent validity have used the infant's observed behavior as the criterion measure. Lounsbury and Bates (1982) found that hunger cries of infants rated as difficult by their mothers were rated as

sounding more "spoiled" and irritating by unrelated mothers than cries of easier infants. Shaefer (1990) gave the RITQ to mothers of 100 babies referred to a crying baby clinic, and found a greater than expected incidence of difficult babies and a smaller than expected incidence of easy babies. One study found differences between criers and non-criers in a laboratory situation to be related to differences on the RITQ mood scale (Roth, Eisenberg & Seu, 1984).

Infants rated as difficult on the RITQ have also been found to sleep less during the night (Weissbluth, 1981), to have difficulty going to sleep (Sanson et al., 1985), to have colic (Sanson et al., 1985), and to be more likely to survive famine conditions (DeVries, 1984). Carey (1985a) found that infants who gained 30% or more of their weight between six and 12 months were overrepresented on the RITQ difficult cluster.

Several studies have examined the relationship between RITQ-rated temperament and the infant's behavior during the Strange Situation procedure (Ainsworth & Wittig, 1969), in an attempt to study the relationship between temperament and attachment. Kemp (1987) found that the infant's scores on the RITQ categories of persistence, mood, and approach/withdrawal were able to discriminate between the infants membership in avoidant, insecure or anxious categories of attachment. Frodi, Bridges and Shonk (1989) also found that RITQ ratings at 4 months were related to the

quality of attachment rated at one year. However, Vaughn, Lefever, Seifer, and Barglow (1989) found that RITQ rated temperament was related to negative emotionality, which is a component of distress during separation, and concluded that temperament measures do not directly predict attachment security but are related via negative emotionality.

Several studies have compared observer ratings of infant behavior with RITQ temperament ratings (i.e. Vaughn et al., 1987; Zeanah, Keener & Anders, 1986), and found significant correlations, though moderate in magnitude. Sameroff, Seifer and Elias (1982) observed infants in the laboratory and in the home setting and attempted to match relevant behaviors with selected scales. They found most relationships to be significant, but reported that the correlations were small in magnitude (none above .26). In their review of temperament instruments, Slabach, Morrow and Wachs (1991) conclude that "if one wishes to study relationships between temperament and ongoing behavior patterns, the strongest instruments appear to be the RITQ (and the BQ)" (page 220).

In summary, research has supported the concurrent validity of the RITQ to some extent, though correlations with criterion measures are generally only moderate. These correlations are similar to those found with other infant temperament measures (Slabach et al., 1991), and not surprisingly low: modest correlations are to be expected in

criterion-related validity studies, as moderating variables most likely play an important role.

Studies assessing the predictive validity of temperament instruments have focused primarily on the relationship between temperament and behavior disorders in children. Thomas and Chess (1968) are commonly cited as demonstrating a significant relationship between infant temperament and behavior problems in childhood. However, the Thomas and Chess study did not find temperament to significantly predict behavior problems until age three. Several authors have reanalyzed the Thomas and Chess data and have found that infant temperament, when considered together with the quality of parenting, did predict behavior problems at age six and beyond (Cameron, 1978). Similarly, using the Infant Characteristics Questionnaire, Bates and Bayles (1988) found that maternally perceived difficult temperament predicted later behavior problems of both an internalizing and externalizing nature. Their measure of adaptability also predicted later problems, particularly in internalizing behaviors. Using the RITQ as a measure of infant temperament, two studies (DiBlasio, Bond, Wasserman, & Creasey, 1988; Sanson, Prior & Oberklaid, 1985) have found a significant relationship between difficult temperament in infancy and behavior problems in early childhood. Nyman (1988), also using the RITQ, found infants who were rated as

difficult in infancy were more likely to be hospitalized later in childhood because of accidents.

In summary, though there have been a limited number studies utilizing the RITQ to predict future behavior on a criterion measure, these studies did support the predictive validity of the RITQ. Research studies using measures comparable to the RITQ have consistently demonstrated that infant temperament, when considered together with other relevant variables, is useful as a predictor of behavior problems later in childhood.

Construct validity. Construct validity focuses primarily on "the test score as a measure of the psychological characteristic of interest" (AERA, APA, & NCME 1985, page 9). Construct validity is typically divided into convergent and discriminant validity. Convergent validity refers to findings that relate the construct under study to other measures that theoretically should be correlated. Discriminant validity refers to findings in which the instrument is found to be related to constructs that are, according to theory, unrelated.

According to the theory underlying the RITQ, temperament is, at least to some extent, biologically based. Research findings demonstrating some heritability or biological correlates of temperament are thus considered to be supporting the instrument's construct validity. Evidence for physiological correlates of temperament has come from

several sources, including twin studies, studies of premature or low birth weight infants, and health related variables.

Two recent studies comparing monozygotic and dizygotic twins have shown significant genetic variance for five and eight of the RITQ scales, respectively (Chen et al., 1990; Cyphers, Phillips, Fulker & Mrazek, 1990), suggesting that monozygotic twins are rated by their mothers as being more similar in temperament than dizygotic twins.

Several studies have examined RITQ rated temperament in premature or very low birthweight infants, and findings have been inconsistent. Medoff-Cooper (1986) found very low birthweight infants to be less adaptable, more intense, and more likely to be rated as difficult. Watt (1987) found small for gestational age infants to be less approaching and more intense, and Spungen and Farran (1986) found high-risk premature infants to be more frequently rated to be more difficult than average. However, other studies have found no significant differences between premature or very low birthweight infants on any of the nine dimensions (Oberklaid, Prior, Nolan & Smith, 1985; Hara et al., 1990). Hara et al. (1990) suggest that the differences found in other studies are a result of their inadequate selection of control groups.

Studies using other physiological variables have found temperament to be related to tonic heart rates (Healy,

1989), respiratory sinus arrhythmia (Richards & Cameron, 1989), recurrence of wheeziness attacks (Priel, Henik, Dekel & Tal, 1990), and cortisol levels during separation (Gunnar, Larson, Hertsgaard, Harris, & Brodersen, 1992).

Thus there is quite a bit of evidence that RITQ rated temperament is related to several physiological variables, suggesting that the RITQ is measuring temperamental traits that are at least in part constitutionally determined.

Evidence for convergent construct validity of the RITQ has been obtained by comparing it with other questionnaire measures of temperament. Support for the instruments construct validity is found if other measures of the same construct are correlated with the instrument. In a recent study investigating the convergent and discriminant validity of temperament measures currently in use, Goldsmith, Rieser-Danner, and Briggs (1991) found "surprisingly strong evidence . . . for convergence among scales intended to measure similar concepts" (p. 556). Specifically, the study found significant and moderately strong (.44 - .74) correlations between RITQ scales and corresponding scales on Rothbart's Infant Behavior Questionnaire and Bate's Infant Characteristics Questionnaire, as rated by both mothers and by teachers. Thus there is evidence to suggest that the RITQ and other questionnaire measures of temperament are measuring similar constructs.

Another method that can be utilized to assess the construct validity of an instrument is the technique of factor analysis. By utilizing factor analytic methods to examine the correlations between the individual items, it is possible to determine which items cluster together as factors, thus assessing statistically whether the questionnaire is indeed assessing one or more cohesive constructs. The RITQ has been factor analyzed in several studies, although most are considered to be weak since the number of questionnaires analyzed was not sufficient to support a valid use of the statistical method (Windle, 1988). One impressive study, however by Sanson et al. (1987) utilizing 2,443 RITQ questionnaires in an Australian sample, found limited empirical support for the nine-dimension structure. They found considerable redundancy in the scales, and concluded that the nine dimensions were intercorrelated, and that many items correlated more highly with separate scales than with their assigned scale. The RITQ scales of Rhythmicity and Persistence were the only scales to emerge from the analysis as measuring relatively pure factors. Based on their results, Prior, Oberklaid, and Sanson (1987) developed the Short Temperament Scale for Infants (STSI), a shortened (30 item) form of the RITQ using only those items that held up as highly correlated with one of the five factors in their model. Their five factors are

labeled Approach, Rhythmicity, Cooperation/Manageability, Activity/Reactivity, and Irritability.

Given these results, and the inconclusive results of other factor analytic studies of the RITQ and other questionnaires based on the nine NYLS dimensions, Slabach et al. (1991) have concluded that "the consistent failure to duplicate the nine-dimensional NYLS model suggests that it may be the theory rather than the instrument that is at fault in this regard" (p.216).

Carey (1989) argues that factor analytic solutions are not always applicable in clinical practice, and that the clinical relevance and usefulness of the NYLS dimensions is sufficient evidence for the validity of the NYLS conceptualization of temperament (Carey, 1985b). A more statistical argument for the usefulness of the NYLS dimensions is provided by Cameron, Rice, Hansen and Rosen (1992). Given the clinical usefulness and practicality of the nine dimensions in their clinical practice, they strove to explain the intercorrelations between the RITQ scales by proposing a causal relationship whereby certain dimensions of temperament affected other dimensions in a predictable fashion. They had found in clinical practice that certain dimensions of temperament (i.e. threshold, intensity) seemed to affect other dimensions (i.e. adaptability, activity) in predictable ways. They developed a causal model based on these relationships, and tested their model statistically,

using a causal analysis. Their results supported the causal relationship between the scales, thus bolstering their argument that the intercorrelations between the nine scales were rational, predictable relationships and thus justified the use of the nine separate dimensions in clinical practice. They comment that temperament traits were never intended to be independent characteristics, and

> temperament trait 'purity' is as welcome as the procrustean bed, forcing the clinician to lop off information revealing the dynamic interplay and causal flow between different aspects of the child's temperament (Cameron et al., 1992, page 3).

In summary, research studies have shown that the RITQ subscales do not hold up to factor analysis; the subscales do not seem to be measuring independent factors. Yet there is some evidence to indicate that although the NYLS dimensions (on which the RITQ subscales are based) are related, the relationship between them is logical and causal. The dimensions of temperament influence each other in predictable ways.

An instrument's discriminant construct validity is supported when the instrument does not correlate highly with measures of constructs that are not theoretically related. Rice and Gaines (1992) have pointed out that most studies have focused on concurrent or convergent validity and more research is needed on the discriminant validity of temperament measures.

However, several studies have demonstrated that RITQ temperament ratings correlate with measures of maternal characteristics or with demographic variables, suggesting that the RITQ lacks discriminant validity. Sameroff et al. (1982) found that socioeconomic status of the mother was related to infant temperament ratings. A similar finding was reported by Brackbill (Brackbill, White, Wilson & Kitch, 1990). Several studies have examined the relationship between maternal characteristics and temperament ratings. RITQ rated temperament has been found to relate to maternal child-rearing attitudes (Frodi et al., 1989), maternal anxiety (Bates & Bayles, 1984; Vaughn, Taraldson, Crichton, & Egeland, 1981; Vaughn et al., 1987; Sameroff et al., 1982), indices of parent mental illness (Sameroff et al., 1982; Affleck, Allen, McGrade & McQueeney, 1983), measures of maternal organization and stimulation (Houldin, 1987), mother-reported family disorganization (Brackbill et al., 1990), and postpartum depression (Cutrona & Troutman, 1986). The most prevalent finding in these studies is that measures of maternal characteristics, especially anxiety, correlate most consistently with the five dimensions used in the determination of difficulty. The dimensions of Mood and Adaptability are particularly influenced by maternal variables (Vaughn et al., 1981).

In a frequently cited and comprehensive study, Sameroff et al. (1982) examined the joint effects of three sets of

variables: maternal socioeconomic status, maternal anxiety, and infant temperament. The study also collected observational data on the infant in the laboratory and at home, and measures of the mother's mental health. They used a hierarchical multiple regression analysis to determine the relative influence of these variables on maternally rated infant temperament, and found that the mother variables explained more of the variance than the child variables, and that maternal characteristics predicted ITQ scores independent of child characteristics (Sameroff et al., 1982). They concluded that

> ITQ scores may be more a result of the projections of the parents than of characteristics of the child...the reduced ability to take perspective found in individuals with mental illness or with the rigid cognitive orientations found in lower SES groups increases the tendency for projected characterizations of the child. In addition, one would expect such projections to be negative, given the emotional distress and economic deprivation of such parents (p.172).

Bates and Bayles (1984) criticize the conclusions of Sameroff et al. (1982) and Vaughn et al. (1981), pointing out that their samples were not representative of the population, including as they did "sociologically extreme groups of mothers, including large proportions of poor, single, and emotionally disturbed women" (Bates & Bayles, 1984, p.113). Bates et al. (1979), using the Infant Characteristics Questionnaire (ICQ), found that maternal social class, personality, and parity correlated with

maternal perceptions of infant difficulty in a moderately sized sample, but that these results were not replicated in a larger sample (Bates, Olson, Pettit & Bayles, 1982).

It could be argued that measures of maternal anxiety obtained concurrently with temperament ratings may be a result of the infant's temperament having an effect on the mother. The mother may be anxious as a result of the stress of dealing with a temperamentally difficult infant. Yet several of the studies finding correlations between maternal variables and RITQ rated temperament were longitudinal (i.e. Brackbill et al.,(1990); Vaughn et al.,(1987); Frodi et al.(1989)), and maternal measures were obtained prenatally, before the infant's behavior could have an effect on the mother. Thus it seems more likely that maternal characteristics are affecting RITQ ratings of infant temperament. Further support for the contention that maternal characteristics influence infant temperament ratings was obtained in two studies which found that mother's RITQ ratings (on the scales of Activity, Rhythmicity, and Mood) obtained prenatally correlated with her RITQ ratings on these scales obtained at six months of age (Zeanah, Keener, Anders, & Baker, 1987; Zeanah, Keener & Anders 1986).

The question can then be raised; "are maternal characteristics affecting infant temperament directly, or are maternal characteristics affecting the <u>rating</u> of infant

temperament via biased maternal perceptions of infants?" The conclusions of Vaughn et al. (1981) have been questioned by raising the possibility that there may be some sort of biochemical mediation between maternal anxiety and infant temperament; thus the infant's temperament may in reality be difficult, at least partly because of the mother's prenatal biochemical influence.

Vaughn et al.(1987) report the results of their study designed to address this possibility. They obtained prenatal placental blood samples of cortisol, adrenocorticotrophic hormone (ACTH), and beta-endorphin, and compared groups of mothers on their levels of these hormones. Though anxious mothers were found to differ from non-anxious mothers on their blood levels of beta-endorphin, there was no difference between mothers of infants classified as difficult or easy on levels of these hormones, suggesting that the placental transmission of these hormones in anxious mothers was not responsible for their infants being rated as difficult.

They also found that prenatal measures of anxiety in the mother were more predictive of RITQ ratings than were measures of observed infant behavior, and concluded that the RITQ lacks discriminant validity in that it seems to be measuring maternal characteristics and not infant temperament.

Carey (1982) has countered these criticisms of the RITQ by pointing out that Vaughn's study had an inadequate match between maternal and professional ratings in content because they were not rating similar behaviors in similar contexts, and that the observations were very time-limited and context-specific (i.e. infants were observed at two feeding times, and at play). Similarly, Bornstein, Gaughran, and Homel (1986) comment that "the strategy of comparing global maternal questionnaires with more focused and delimited observer measures seems inappropriate...the convergent validity of parent reports cannot be properly evaluated by comparing sets of observations obtained by different methods and with different demands" (Bornstein et al., 1986, p. 188).

Bates and Bayles (1984) have pointed out that several studies have demonstrated that mothers can be trained to report behavior as accurately as objective observers, when criteria are made explicit and specific behaviors are outlined for observation (i.e. Cummings & Radke-Yarrow, 1981). He concludes that the vast amount of information available by maternal report should not be discarded. Instead, researchers would benefit by studying sources of maternal bias and taking these into account when designing and utilizing maternal ratings of temperament. Of note is the finding by Power, Gershenhorn and Stafford (1990) that the RITQ, which asks mothers to rate the frequency of

specific behaviors, correlated more highly with variables reflecting infant behavior than did the ICQ, which asks mothers to rate their infant's behavior in less specific terms. This result suggests that maternal ratings of specific patterns of infant behavior relevant to temperament may be less biased by maternal characteristics than are maternally rated perceptions of infant temperament.

Rationale and Description of the Current Study

The prevalence of studies finding correlations between maternal characteristics and maternally rated infant temperament have led some researchers to conclude that mothers are not accurate and objective reporters of their infant's behavior and therefore maternal ratings of temperament should be avoided. Still others have countered that despite their bias, mothers possess a tremendous volume of experience with their infants that cannot be matched by limited, though objective laboratory observation.

One method of obtaining relatively objective behavioral data as it occurs in the home environment over a representative period of time is to utilize structured diary reports. This method requires mothers to record their infant's behavior as it occurs over several days. Specific behavioral criteria are used to minimize potential maternal bias, and diary content is matched to questionnaire content quite easily. Despite the potential value of such a method

in obtaining extensive information about an infant's behavior in a natural environment, no studies to date have used diary reports to assess infant temperament. Accordingly, the present study aimed to provide validity data on the RITQ by using the comparison measure of maternal diary reports.

A diary report form was designed to match in content items on the RITQ, using specific behavioral criteria to rate and record infant behavior as it occurred over a four day period. This study limited itself to examining the five dimensions involved in Thomas and Chess's conceptualization of difficulty (Approach, Adaptability, Mood, Intensity, and Rhythmicity). These dimensions were chosen because they have proven to be the most useful in clinical practice and the most salient in research to date. Additionally, to attempt to measure all nine NYLS dimensions in diary form would most likely prove to be too laborious for most mothers.

In order to further assess the construct validity of the RITQ and to shed additional light on possible discrepancies between maternal ratings of temperament and diary recorded infant behavior, two additional measures were included in the study for comparison. Mothers were asked to respond to Bates' Infant Characteristics Questionnaire (ICQ), a factor-analytically derived measure that is also based on the five NYLS dimensions involved in the perception

of difficulty. This measure was chosen because of its sound psychometric properties and because it was specifically designed to tap more global maternal perceptions of infant temperament, which according to research results to date should be more distantly related to objectively observed infant behavior than ratings of behavior in specific contexts. An observer rating form was also designed to match the content of RITQ scales, and ratings were obtained on two occasions to provide objective observational data for comparison.

Specific Hypotheses

1.) It was expected that scale scores on the five RITQ dimensions involved in the assessment of difficulty (Adaptability, Intensity, Approach/Withdrawal, Mood, and Rhythmicity) would correlate at a moderate level with corresponding Diary scale scores.

2.) Diary scale scores were also expected to correlate at a low to moderate level with corresponding ICQ scores measuring the same dimensions. These correlations were not expected to be as strong as those obtained between the Diary and the RITQ, because the ICQ was designed to measure maternal perceptions rather than actual infant behavior.

3.) It was expected that observer rating scores on each of the dimensions would correlate with corresponding scale scores on the RITQ and the ICQ, though correlations were expected to be low, because these ratings are based on

very time-limited and thus less representative samples of behavior. It was also expected that observer rating scores would correlate with Diary scores. A moderate correlation was expected, since both observer and Diary scores are assumed to be relatively objective measures based on infant behavior and less subject to maternal bias than are the questionnaire measures.

RESEARCH METHODOLOGY

Subjects

Subjects were chosen from the population of white married parents of eight month old infants. This homogenous population was chosen in order to control for the variables of infant age, race, and marital status of the mother. The sample chosen was a convenience sample; subjects were recruited on a voluntary basis from lists of potential subjects who had previously volunteered for a research study, from birth announcements in the local paper, and from word of mouth referrals.

Potential subjects were contacted by phone and asked to participate in a study of infant temperament, in which they would be interviewed in their home; their baby would be observed, and they would be asked to fill out several questionnaires which rated and recorded their baby's behavior. The Diary was not specifically mentioned until after the RITQ was administered, to prevent possible bias in rating and selective memory of specific item responses.

The completed sample consisted of 45 mother-infant

pairs. As selected, the mothers and infants were all white and the mothers were married. The infants were an average of 8 months old (range 7 to 9 months), 71% were female, and 71% had no siblings. The mothers' mean age was 30 (range 19-40), and they had an average of 16 years of education (range 12-20). Fathers' mean age was 33 (range 21-44).

Instruments

The instruments used in the study were an interview schedule, the Revised Infant Temperament Questionnaire (RITQ), the Infant Characteristics Questionnaire (ICQ), an Observer Rating Form, and the Diary.

Interview schedule. The interview schedule was designed to gather demographic information (ages, occupations, education, siblings of the infant), situational information (work schedules, child care arrangements, how long married, husband's participation, experience with babies, size of home and ownership status), and attitudes and opinions (perception of infant's personality, opinions/behavior regarding schedules) of the mother. (See Appendix.)

RITQ. A 54-item form of the RITQ was administered, which was revised from the original 95 item RITQ by deleting the questions assessing threshold, persistence, distractibility, and activity, leaving only the dimensions involved in the designation of an infant as easy, intermediate or difficult. The original instructions

accompanied the questionnaire. (See Table 1 for reliability data.)

ICQ. The Infant Characteristics Questionnaire (ICQ) (Bates, Freeland & Lounsbury, 1979), a 27-item questionnaire assessing maternal perceptions of temperament, was administered following the RITQ. The ICQ was developed as a brief screening device for assessing maternally perceived difficultness in infancy. The ICQ's 27 items separate into four factor-analytically derived dimensions: Fussy-Difficult (nine items), Unadaptable (five items), Dull (four items) and Unpredictable (three items). Bates et al. (1979) report internal consistency estimates of .79, .75, .39 and .50, and test-retest reliability coefficients of .70, .54, .57 and .47, respectively, for the four factors.

Observer rating form. Behavioral ratings were devised to measure the infant's initial response to the interviewer, adaptability to the interviewer's presence, mood rating while playing with the interviewer, approach/withdrawal response to the novel toy, adaptability to the toy, and mood while playing with the toy. The observer rating form was refined during the interviewing of 10 pilot subjects; interrater agreement was 90%, and no discrepancies in ratings exceeded one scale point. The second observation measured these same items again (a different toy was introduced). Space was alloted for comments, observer description of baby and observer impressions, including note

of the mother's comments as to how typical her baby's behavior was during this time, and factors that may have affected behavior (i.e. teething, illness, fatigue). (See Appendix.)

Diary. The diary was a 10-page handout which included one page of instructions and one page describing a mood scale which was to be used to rate the baby's behavior. The mood scale was a seven-point behavior rating scale with each point describing an infant's specific behaviors expressing mood. Mothers were instructed to post the scale in a prominent place for repeated reference while filling out the diary. Of the remaining eight pages, two pages were alloted for each of four days of diary recording. The first page of each day asked five specific questions, which asked the mother to introduce something new each day (e.g. new vegetable or new napping place) and record the baby's reaction. Nighttime awakenings, responses to new people and new situations during the course of the day were also questioned and rated. The final questions asked how typical the infant's behavior was that day, and asked for comment on any factors that may have affected the baby's behavior. The second page of each day was constructed in a chart format, to record the time of each feeding, each nap, each bowel movement and each bath, and to record the baby's mood rating at each of these intervals. Mothers were also encouraged to write comments freely on the back of each page or in the

margins, especially if the elicited questions missed important information. (See Appendix.)

Procedure

Once potential subjects agreed to participate, an interview time was set up when the baby would be awake and fed, and a researcher visited the home. Written informed consent was obtained, and a brief structured interview followed. The mother was then given the two infant temperament questionnaires (the RITQ and the ICQ) to fill out, while the interviewer played with the baby on the floor in the same room. After 15 minutes of interaction with the baby, a toy was introduced to the infant by the observer. Toys were chosen to be large, colorful and attractive to an eight-month old infant. Interviewers (other than the author) were trained by the author to observe the infant's reactions and behavior, and were instructed to record their observations and ratings shortly after leaving the home. Each interviewer was accompanied by the author on their first two interviews; the first time simply observing and rating the infant's behavior and the second time being observed by the author.

After the mother finished the questionnaires, the diary was introduced, with the explanation that a more complete picture of the baby's temperament could be obtained with a recording of his/her behavior over several days (see Appendix: Instructions for introduction of diary). Specific items were explained, and mothers were told they would be called within

a few days; phone numbers of the interviewer and of the author were given in the event any questions or concerns should arise. They were also told that when they finished the diary, the interviewer or the author would again visit to pick it up, observe the baby again, and provide them with verbal feedback on their responses to the temperament questionnaires. Specific questions about the purpose of the study were answered by saying that this was a study of the measurement of infant temperament, not their child individually, and that if they were interested, the specific purpose of the study would be explained in more detail upon completion of the diary.

Design and Data Analysis

Description of Obtained Scores

RITQ. The 54-item RITQ yielded five scores, one on each dimension of temperament: Mood, Intensity, Approach/Withdrawal, Adaptability, and Rhythmicity. Additional scores were obtained by summing scores on individual items within each dimension assessing behavior in a specific context (i.e. bathtime, feeding, sleeping).

Observer ratings. The observer rating form yielded scores on four of the five dimensions. Rhythmicity could not be measured in two observations.

Diary. The Diary yielded scores on each of the five dimensions, derived as follows:

Rhythmicity: Variability (as measured by the standard deviation in time across days) in first wake time, first nap

time, length of first nap, second nap time, length of second nap, bedtime, first feeding time, last feeding time, and first bowel movement time were obtained and summed to provide a total score for rhythmicity. Separate scores for regularity of sleep, feeding and bowel movements were obtained by summing the scores within these content areas.

Mood: Separate mood ratings for each interval were averaged to provide a mean mood score. Mood scores for each context were thus obtained by summing the mood scores pertaining to sleep, feeding, diapering, and bathtime.

Intensity: A rating for intensity of response was obtained by taking each diary rating (on a scale of 1-7) and calculating the distance from the mid-point of 4=no response, and adding this to the base score of 4, resulting in intensity scores with a range of 4-7. This derivation was based on the assumption that more intense reactions, whether positive or negative, according to theory will deviate more from the middle or "no-response" rating. Thereby a Diary rating of 1 equals in intensity a mood rating of 7, a rating of 2 equals in intensity a rating of 6, and a rating of 3 equals in intensity a rating of 5.

Approach/Withdrawal: Approach/Withdrawal scores were obtained by averaging ratings on the questions eliciting the infant's first response to novel stimuli (vegetable, sleeping place, people, and situations).

Adaptability: Adaptability scores were obtained by averaging ratings on items designed to elicit the infants "eventual" response to new situations (new vegetable, second try; new nap place, second try; new situations, "later on" ratings).

Statistical Analysis

Primary analyses

To test the first hypothesis, that Diary subscale scores would correlate with RITQ subscale scores, correlation coeffiecients were obtained between the five separate RITQ scale scores and the five separate diary scale scores.

To test the second hypothesis, correlation coeffiecients were obtained between the Diary subscales, the RITQ subscales, and the ICQ factors representing the dimensions measured for each subscale: RITQ and Diary Mood with ICQ Fussy-Difficult; RITQ and Diary Approach and Adaptability with IBQ Unadaptable; RITQ and Diary Rhythmicity with IBQ Unpredictable, and RITQ and Diary Intensity with IBQ Dull.

To test the third hypothesis, the observer rating scores were compared and correlations obtained with corresponding RITQ scale scores, IBQ factor scores, and diary scale scores.

Supplemental analyses

In order to assess whether correlations between the Diary and the RITQ were affected by varying emphasis on situational context, subscores measuring each specific context within each dimension were compared. For example, the mean ratings on questions pertaining to regularity in sleep patterns on the RITQ were correlated with the sleep rhythmicity score on the Diary. To more closely analyze the data, those specific Diary items matching RITQ items almost exactly in content were also compared.

To shed more light on the meaning of the lack of correlation between the Diary Rhythmicity score and the RHythmicity scores on the RITQ, mothers were divided into two groups and compared based on their reported scheduling practices and opinions regarding scheduling. These two groups were then compared using ANOVA, testing the hypothesis that more "irregular" mothers would have higher Diary Rhythmicity scale scores.

In order to assess whether the correlations between the RITQ scales and other scales were due in part to the poor factor structure of the RITQ, factors derived from the RITQ items reported by Sanson et al. (1987) and used in the STSI (Prior, Oberklaid & Sanson, 1987) were utilized for comparison with Diary scale scores and ICQ factor scores.

RESULTS

Descriptive Statistics

Responses to Interview Schedule

Most of the mothers stayed at home most of the time: 40% did not work outside the home; 45% worked part-time and 15% worked full-time. According to mothers' report, 76% of the fathers participated more than minimally in their infants' care. All of the fathers worked at least 30 hours per week outside the home, averaging 45 hours per week. Most of the parents of the infants owned (were buying) the home they were living in (78%). The mothers and fathers had been living together an average of 6 years (range 2-17). All mothers reported their infants to be in generally good health: 91% reported one or less physicians' visits for illness in last two months. Most of the mothers in the sample had prior experience with infants; only 16% claimed to have no experience with infants prior to their child's birth, while 48% claimed to have a great deal of experience. When asked to describe their infants as newborns, 18% stated that their

infants were difficult in the first few months, while 40% described them as easy to care for as newborns.

When asked if they followed a feeding schedule, 48% of the mothers stated that they followed a schedule all or most of the time. However, only 30% of the mothers reported that they put their infant to sleep at scheduled times. When asked about how they felt about schedules in general, 35% of the mothers stated that they did not like schedules, while 56% felt that schedules were important as long as they were flexible.

Descriptive statistics on Diary, RITQ and ICQ

Table 2 shows the sample means and standard deviations of dimension scores for the RITQ, the ICQ and the Diary, and internal consistency estimates for the RITQ and ICQ. The table also includes means and standard deviations of each questionnaire's standardization sample. Means and standard deviations obtained in the current sample matched very closely the standardization samples of both the RITQ and the ICQ. Internal consistency estimates (utilizing Cronbach's alpha) for the RITQ were low to moderate, while estimates for the ICQ were moderate (given the small number of items in three of the scales).

Despite the present sample means and standard deviations being similar, however, when the current sample was categorized according to RITQ instructions into Easy, Difficult and Intermediate categories, none of the infants was

Table 2.

Sample Descriptive Statistics for the Diary, RITQ and ICQ

Scale	Sample			Standardization Sample	
	Mean	S.D.	I.C.*	Mean	S.D.
RITQ					
Rhythmicity	2.32	.56	.68	2.36	.68
Intensity	3.40	.61	.40	3.42	.71
Mood	2.72	.61	.59	2.81	.68
Adaptability	2.03	.46	.46	2.02	.59
Approach/ Withdrawal	2.37	.55	.57	2.27	.78
ICQ					
Fussy	16.78	5.47	.78	17.77	5.88
Unadaptable	8.04	2.92	.51	8.90	4.00
Dull	6.11	2.80	.58	5.88	1.85
Unpredictable	7.29	2.68	.53	7.32	2.69
Diary					
Rhythmicity	13.93	4.82			
Intensity	28.33	1.85			
Mood	2.87	.56			
Adaptability	2.79	.97			
Approach/ Withdrawal	2.78	.80			

* I.C. = Internal consistency

classified as difficult. Easy infants made up 38% of the sample, and the remainder were classified as intermediate. This categorization changed only slightly when sample means were utilized instead of Carey's norms as classification criteria. The classification of one infant changed from intermediate to difficult.

Results of Hypotheses Testing

Relationship between scale scores of the Diary and the RITQ

To test the first hypothesis, that Diary scores would match corresponding RITQ scores, correlation coefficients comparing the Diary scale scores with the RITQ scale scores were obtained. Only two of the Diary subscale scores correlated at a statistically significant level with their corresponding RITQ scale scores (see Table 3). As expected, there was a moderate correlation between Diary Mood and RITQ Mood ($r = .47$, $p < .01$) and between Diary Intensity and RITQ Intensity ($r = .34$, $p < .05$). To rule out the possibility of non-linear relationships between variables or extreme outliers affecting the overall correlation, plots of the relationship between each of the Diary scales and corresponding RITQ scales were obtained and scrutinized. No apparent non-linear relationships were observed, and no extreme outliers were found.

Relationship between ICQ factors and corresponding scales on the RITQ and the Diary

To test the second hypothesis, correlations were obtained comparing the four ICQ factor scores with corresponding scale scores on the Diary and the RITQ. It was expected that ICQ factor scores would correlate at a low to moderate level with corresponding scale scores on the Diary. This hypothesis was partially supported. The ICQ factors correlated at a significant level with all five corresponding RITQ scales, but with only three of the five Diary scales (see Table 3).

Table 3.

Correlations Between Diary, RITQ and ICQ Dimension Scores

Dimension	Diary/RITQ	Diary/ICQ	RITQ/ICQ
Approach/ Withdrawal	.11	.30 *	.50 ***
Adaptability	-.08	.31 *	.33 *
Mood	.47 **	.46 **	.53 ***
Rhythmicity	.09	.06	.30 *
Intensity	.34 *	-.26	-.40 **

* p<.05
** p<.01
*** p<.001

It is interesting to note that two of the diary scales demonstrating no significant correlation with matching RITQ scales were related to the corresponding ICQ factor score. The correlations between both Diary scales of Approach/Withdrawal and Adaptability, and ICQ Unadaptable were significant and moderate ($r=.30$ and .31, respectively, $p<.05$). The Diary mood subscale scores correlated with the ICQ Fussy-Difficult scale scores ($r=.46$, $p<.01$). as well as with the RITQ Mood scale scores. The correlation between Diary intensity and ICQ Dull ($r=-.26$, $p=.08$) approached significance. The negative correlations with the Dull factor were expected, since high scores on the factor "Dull" apparently indicate lower intensity.

Comparison of Observer's Rating Scores with Other Measures

To test the third hypothesis, the four observer rating scores (for approach, adaptability, mood and intensity) were compared with corresponding scale scores on the Diary, the RITQ and the ICQ (See Table 4.). No significant correlations were found between any of the observer scores and other scores, although the correlations between observer scores and Diary scale scores came close to reaching significance (Approach: $r=-.24$, $p=.12$; Adaptability $r=.30$, $p=.052$; Mood: $r=.28$, $p=.06$; and Intensity: $r=.21$, $p=.16$).

Table 4.

Correlations Between Observer Scale Scores and Corresponding Diary, RITQ and ICQ Scores

	Observer Dimensions			
	Approach	Adaptability	Mood	Intensity
Diary	.24	.30	.28	.21
RITQ	.18	.20	.06	.04
ICQ	.12	.11	.08	.007

Supplementary Analyses

Sub-analyses by context

In order to more fully understand and explain the obtained correlations between subscale scores on the Diary and on the RITQ, RITQ and Diary items within subscales were divided and grouped by context and compared. Diary measures specifically matching RITQ individual items in content were also compared.

Mood. Within the Mood subscale, items were divided according to whether their contexts involved sleep/waking, feeding, or diapering. RITQ items on the Mood scale with a sleep/wake context were significantly correlated with items on the Diary involving mood related to sleep/waking ($r=.31$, $p<.05$)(See Table 5.). A similar correlation was obtained for feeding situations ($r=.34$, $p<.05$). However, the correlation between the RITQ and the Diary items rating mood

Table 5.

Mood Subscale Comparisons by Context

Context	Diary/RITQ Correlation
Sleep/Waking	.31 *
Feeding	.34 *
Diapering	.23

* p<.05

during diapering did not reach statistical significance (r=.23, p=.13).

Table 6.

Approach/Withdrawal and Adaptability Context Comparisons

Context	RITQ/Diary Approach	RITQ/Diary Adaptability
Feeding	.24	.31 *
New nap place	.46 **	.35 *
New situations	.12	.16
New people	.14	N/A

* p<.05
** p<.01

Approach/Withdrawal and Adaptability. Within the RITQ and Diary subscales representing the Approach/Withdrawal dimension, separate correlations were obtained for approach/withdrawal ratings in the contexts of new

situations, new people, introduction of a new vegetable, and a new sleeping place. The only correlation between the RITQ items and their matching Diary items to reach significance was Approach to new napping place ($r=.46$, $p<.01$)(See Table 6.).

Within the Adaptability scales, similar results were found for adaptability to new napping place ($r=.35$, $p<.05$). Dairy-reported adaptability to a new food also matched mother's report of eventual acceptance of change in food on the RITQ ($r=.31$, $p<.05$). The mean of the four Diary-reported adaptability to new situations ratings did not correlate at a significant level with the mean of the RITQ items measuring adaptability to new situations.

Intensity. RITQ items and subgroups of items matched in content to Diary ratings of intensity were compared; only one context comparison (measuring intensity of response during diapering) even approached significance ($r=.26$, $p=.08$). None of the other expected relationships between RITQ and Diary subgroups of items measuring intensity in specific contexts reached statistical significance (See Table 7.), despite the fact that a significant correlation between Diary intensity score and RITQ intensity score had previously been obtained (See Table 3).

To examine further the correlation between Diary intensity and the ICQ Dull factor, correlations were obtained between total intensity score on the Diary and the

Table 7.

Intensity Context and Item Comparisons

Context	RITQ/Diary Correlation
Feeding	.15
Diapering	.26
Bath	.07
Total score	.34 *

ICQ Items	Diary/ICQ Correlation
15. how active in general	-.02
16. Smile/happy sounds	-.24
23. How excited when play	-.32 *
Total Score	-.26

* $p<.05$

three individual ICQ items making up the ICQ Dull factor. Only one of the three items on the ICQ dull scale correlated at a significant level with the Diary intensity score. This item was the only one measuring intensity of response (How excited does your baby become when people play with or talk to him/her?). The other two items appear to relate more to mood and activity (How much does your baby smile and make happy sounds? and How active is your baby in general?).

Rhythmicity. Items on the RITQ Rhythmicity scale matched in content to specific rhythmicity measures on the diary were compared and correlations obtained. Within the sleep context, the only expected relationship to even

Table 8.

Rhythmicity Item Comparisons By Context

Context	Diary/RITQ Correlation
Sleep:	
A.M wake time	.25 +
first nap time	-.02
length of naps	.23
P.M. bed time	.13
Feeding:	
wants meals same time	-.09
Diapering:	
bowel movements same time	.31 *

+ p <.10
* p <.05

approach significance was regularity of time of waking in the morning, with a correlation of .25 (p=.097) (See Table 8.). Within the feeding context, one item matched most closely in content to the Diary measure (The infant wants and takes solid food feedings at about the same time each day) was compared with the Diary rhythmicity subscore pertaining to regularity in time of feeding; this correlation was small and nonsignificant. Diary measured rhythmicity of bowel movements (measured by variance in time of first bowel movement each day) did correlate significantly with mother's response to the matching item on the RITQ (r=.31, p=.03).

Analysis of scheduling opinions/practices

In order to further understand the relationship between questionnaire-reported rhythmicity and Diary-reported rhythmicity, and to determine whether scheduling practices and opinions about scheduling practices had any effect on Diary-reported rhythmicity, subjects were compared based on their responses to the interview questions about scheduling. First, subjects who stated that they followed a feeding schedule most of the time (48%) were compared with subjects who did not follow a regular feeding schedule. There was no significant difference between the two groups in variance of feeding times on the Diary. Next, subjects who stated that they usually put their infant to bed at regularly scheduled times (30%) were compared with subjects who reported that they did not follow a sleep schedule. These two groups did not differ in their Diary sleep rhythmicity scores. Subjects were then categorized into two groups, based on their response to the question "How do you feel about schedules in general?" Diary and RITQ rhythmicity scores for those who did not like schedules (35%) were compared with those who beleived schedules were important. The two groups did not differ in their mean responses to the RITQ rhythmicity scale, but subjects who did not like schedules had higher Diary rhythmicity scores ($F(1,42)=9.87$, $p<.003$), meaning their infants demonstrated more variability in time

of sleep, feeding and bowel movements over the four days reported in the Diary.

Comparisons using STSI factors

In order to shed additional light on whether the low correlations between RITQ scales and Diary scales were due primarily to the psychometric inadequacy of the RITQ, comparisons were made utilizing Sanson et al.'s (1987) factor-analytically derived scales of the RITQ. Only the first two factors could be utilized, since the other three factors included items from scales not included in this study. These first two factors (Approach/Adaptability and Rhythmicity) were compared with corresponding Diary scales (Approach and Adaptability, Rhythmicity). The correlations obtained were very small and nonsignificant. However, these STSI factor scores did correlate well with corresponding ICQ factor scores (Approach/Adaptability with Unadaptable: $r=.61$, $p=.0001$; Rhythmicity with Unpredictable: $r=.34$, $p=.02$).

SUMMARY AND CONCLUSIONS

Summary

This study was designed to assess the construct validity of the Revised Infant Temperament Questionnaire (Carey & McDevitt, 1978) utilizing the comparison measure of Diary reports. Previous studies have questioned the construct validity of the RITQ because it has been shown to correlate more highly with maternal characteristics than with laboratory observations of infant behavior. Supporters of the RITQ have countered that these observations were based on limited samples of behavior not matched in content to the RITQ, and that the RITQ does in fact measure infant behavior reflecting temperament and not simply biased maternal perceptions of infants.

Diaries were designed to match the content of the RITQ as closely as possible, and were filled out by 45 white, married mothers of eight month old infants over four days. This study was limited to the five dimensions (and corresponding RITQ scales) involved in the assessment of difficultness. Observer ratings obtained during two home

visits and the Infant Characteristics Questionnaire (ICQ) were also used for comparison.

It was hypothesized that RITQ scale scores would yield correlations of a moderate size with corresponding Diary scores, supporting the contention that RITQ ratings reflect actual infant behavior, and not simply maternal characteristics. The ICQ was designed to tap more global maternal perceptions and is worded in much less specific and contextual terms; hence it was expected that this measure would also correlate with the Diary, but at a low to moderate level. It was expected that the observer measures would correlate at a moderate level with the Diary measures and to a lesser extent with the RITQ and the ICQ.

These hypotheses were supported only in part. Two of the Dairy scales (Mood and Intensity) did correlate at a significant level with corresponding RITQ scales. However, three of the five Diary scales (Approach, Adaptability, and Rhythmicity) had no relationship to corresponding RITQ scales. It was also found that twice as many of the Diary scale scores correlated at a significant level with corresponding ICQ factor scores than with RITQ scale scores, despite the fact that the Diary was specifically designed to match RITQ content. The Diary measure of rhythmicity had no relationship to either RITQ Rhythmicity or ICQ Unpredictability. Observer rating scores did not correlate at a significant level with their corresponding scores on

the RITQ, the ICQ, or the Diary, although correlations between the Diary and the observer scores approached significance.

Additional analyses were performed to determine the effect of maternal scheduling practices and opinions on mother's ratings of infants' rhythmicity on the Diary. Although mothers' report of scheduling practices did not correlate with the degree of infants' rhythmicity on the diary, her opinions about these practices did: mothers who stated that they did not like schedules had infants who obtained higher (more variable) rhythmicity scores on the Diary, suggesting that their babies were less rhythmic.

Additional analyses were also performed to determine if the weak pattern of correlations between the Diary scales and the RITQ scales were due primarily to the RITQ's weak factor structure. Two factor analytically derived dimensions from the RITQ used in the Short Temperament Scale for Infants (Prior et al., 1987) were compared with corresponding Diary scales (Approach and Adaptability, Rhythmicity): correlations were low and nonsignificant.

Discussion

Sample Characteristics

The sample utilized in this study was limited to white, married, and educated mothers. This sample is similar demographically to the standardization sample for the RITQ, and sample means and standard deviations were also similar,

suggesting that the present sample was appropriate for the designed use of the instrument. The proportion of difficult infants was much smaller in the current sample: this may be due to the sample size being smaller (45 versus the standardization sample of 200), or it may be that the current sample contained more infants classified as easy. Given the characteristics of this sample, the high proportion of easy infants is not surprising. This finding is consistent with previous studies which have found a disproportionate number of difficult infants in lower SES, nonwhite populations (i.e. Brackbill et al., 1990; Sameroff et al., 1982). Generalizations of the results of this study must thus be limited to white infants of educated mothers in intact families.

Relationship Between the Diary, the RITQ and the ICQ

The results of the comparisons between Diary scales and questionnaire scales will be discussed within each dimension of temperament measured.

Approach and Adaptability. These two dimensions will be considered together in this discussion because results for the two scales were very similar, they are highly correlated, and factor analytic studies have consistently found them clustered together. It was hypothesized that Diary and RITQ scales measuring approach and adaptability would be correlated. This result did not occur, although specific items on the RITQ rating approach and adaptability

to new food and new napping place did correlate with their corresponding measures on the Diary. Additionally, both Diary scales did correlate at a significant level with the ICQ scale of Unadaptable.

These results are most likely attributable to several factors. First, the Diary scales of Approach and Adaptability are composed of relatively few samples of behavior (10 and 6, respectively), thus increasing the potential sampling error and decreasing the power of any comparisons using this measure. Yet both of these scales did correlate with the ICQ's corresponding measure (Unadaptable), suggesting that the low number of items on the Diary is not the only factor contributing to the obtained results. It is possible that the RITQ scales' poor psychometric properties also contributed to the lack of significant results. Given the low to moderate internal consistency on these measures, the RITQ items within these scales may not be measuring one cohesive factor. Some of the RITQ items themselves may be at fault. New napping place and new food were the only RITQ items to correlate well with their corresponding Diary measures. Hence it may be that approach and adaptability are so situation-specific that only very specific questions (i.e. approach to new vegetable) elicit reliable responses. An infant's response to new people or new situations is likely to be highly variable and dependent on the particular person or situation

in question. This could be one reason why the ICQ seems to measure the construct of Approach/Adaptability most adequately. These items ask mothers to rate how their infant "typically" responds to a new person or situation, rather than specifying the context (Doctor, babysitter, store, new child), as the RITQ does. The RITQ items on these scales may not provide a representative sample of behavior reflecting the temperament dimension of approach/adaptability.

Mood. As hypothesized, the Mood scale on the Diary correlated at a relatively high level with the RITQ Mood scale, supporting the validity of this RITQ scale. The relatively high degree of correlation obtained may be attributable in part to the large number of samples of behavior in which mood was rated on the Diary (66), thus increasing the scale's reliability. The correlation between the Diary Mood scale and the corresponding scale of the ICQ (Fussy-Difficult) was also relatively high (see Table 3). Additionally, all of the Diary Mood ratings correlated at a significant level with the corresponding RITQ items and groups of items, regardless of context, suggesting that the mothers' rating of mood does not seem to be situation-dependent.

One explanation of the consistent correlations between the measures is that the similarity between the three measures (the Diary, the RITQ and the ICQ) in the task of

rating the positiveness or negativeness of their infant's mood contributed to the results. The Diary measure was designed to be more objective, in that mothers were asked to look at their infant on each of 66 occasions and rate his/her behavior on a carefully derived behaviorally indexed scale. Yet the element of subjective rating remained present, thus allowing for some degree of maternal bias. This maternal bias is more obviously present in the questionnaire measures, which ask for more global assessments of the infant's predominant mood in general contexts.

Research has shown that the temperamental dimension of Mood is the dimension most subject to maternal bias and preconceptions (Bates, 1980). It has consistently emerged as one of the dimensions correlating with maternal characteristics, and is the most salient feature of the concept of difficultness, which some researchers view primarily as a social perception. Zeanah et al. (1986) found that mood was one of the three maternally rated RITQ dimensions to remain stable from prenatal to six months postnatal measures, and the only one that could not reasonably be tied to perception of fetal movements. This consistent research finding (that Mood is the dimension most strongly related to maternal characteristics) supports the contention that all three measures in the present study were strongly influenced by maternal variables, hence increasing

the common variance between the measures and thereby increasing the correlation.

Intensity. As expected, the Diary measure of Intensity correlated at a significant level with the RITQ Intensity score, and to a lesser extent with the ICQ Dull factor. What is interesting, however, is that none of the individual context correlations were significant: i.e. intensity of response rated during feeding was not related to responses to RITQ Intensity scale items in feeding contexts. This result may reflect the disparity in measurement methods between the Diary and the RITQ, in that the RITQ asks mothers to rate the intensity of the infant's response in specific contexts, whereas the Diary measure of intensity was obtained less directly, being derived (as the extremity of response) from mood, adaptability and approach ratings. Alternatively, this may suggest that the correlations are due more to the mothers' tendency to rate at extremes than to the infants' actual intensity of response in specific contexts. If this were the case, however, one would expect the Diary Intensity score (which is essentially a measure of extremeness of rating) to correlate equally well with all RITQ and ICQ scales. This was not the case, which suggests that although maternal tendency to rate at extremes may be a factor in the obtained results, it is not the only factor accounting for the correlation between Diary and RITQ Intensity.

The ICQ Dull factor is composed of three items, only one of which seems to measure intensity of response (the other two are related to mood and activity). The Diary measure of intensity correlated significantly with only this item on the ICQ, which was not surprising. This finding also supports the contention that the Diary measure of Intensity is indeed measuring intensity of response.

The significant correlation between Diary rated Intensity and RITQ Intensity score supports the construct validity of this RITQ scale. The measures correlated despite being derived in very different ways. There is little evidence to suggest that maternal bias was a significant influence in these results. This is not surprising, since RITQ rated Intensity has not been found to be a variable that is highly correlated with maternal characteristics or behaviors.

Rhythmicity. Diary measured rhythmicity did not correlate with either the RITQ Rhythmicity scale or the ICQ Unpredictable scale; this result was not expected. There are several reasons why this lack of relationship may have occurred. First, the Diary measure of variability among four days may not have been an adequately large sample to accurately measure rhythmicity. These four days were not required to be consecutive days, which may have increased the error variance (i.e. a baby's schedule may have been gradually shifting earlier or later, while remaining fairly

regular, and skipping days increased this variance). A second possibility is that the Diary rhythmicity score did not reflect the infant's rhythms directly because the mother and the environment have such a profound influence on daily scheduling. An infant may be hungry or sleepy at regular times, but he or she may not be given the opportunity to eat or sleep at these times. In this case, the mother may rate the baby as rhythmic on the RITQ and ICQ, but their experience together recorded as Diary events reflect her patterns, not her infant's (or some interaction of the two), thus decreasing the correlation between these measures. Two findings in this study support this conclusion. First, the only individual RITQ items to correlate with corresponding Diary items were those measuring time of waking in the morning and time of first bowel movement each day. Of all the rhythmicity variables utilized, these two are the least subject to maternal influence. A mother has much less control over her infant's morning wake time or bowel movement time than she does other variables, i.e. time of first nap, or first feeding time. Second, when mothers who stated they did not like schedules were compared with mothers who did like schedules, the mothers who did not like schedules had higher Diary rhythmicity scores (despite not having higher RITQ Rhythmicity scores), suggesting that their opinions were reflected in their behavior. Irregular mothers did indeed have more irregular schedules and more

variance in sleep and feeding times. Thus it may be that the Diary measured the rhythmicity of the mother as much or more than the rhythmicity of the infant. Other studies have also found maternal and family characteristics to have an influence on mother-reported rhythmicity of her infant. Brackbill et al. (1990) found that mothers who obtained scores reflecting family stability on the Family Dynamics Measure rated their infants as more rhythmic on the RITQ than mothers who obtained scores reflecting family disorganization. A similar finding was reported by Tomasdottir et al. (1991), with an Icelandic sample. Sprunger, Boyce and Gaines (1985) also found that infant rhythmicity was related to family rhythmicity. They reported that infant rhythmicity (as reported on the Perceptions of Baby Temperament Scale) was significantly correlated with the mothers' ratings on a measure of family routinization. These results suggest that it is likely to be very difficult to design an objective measure of an infant's innate tendency to be rhythmic without extensive environmental manipulation.

In summary, the hypothesis that RITQ scores would correlate with the comparison measure of Diary reports was only partially supported. Only two of the five scales on the RITQ correlated at a significant level with corresponding Diary scale scores. Additionally, there is some evidence to suggest that at least one of these

correlations (Mood) may be attributable in part to the fact that this diary scale, like the RITQ, was derived from relatively subjective maternal ratings, which necessarily involve a certain degree of bias. Thus the correlation obtained here may be due in part to the fact that all three Mood measures (the RITQ, the Diary and the ICQ) are measuring primarily maternal perceptions. The Diary scale of Intensity is less subject to this bias, yet this scale correlated at a significant level with the RITQ and at a near-significant level with the ICQ. This finding supports the validity of the RITQ scale of Intensity. The only Diary scale apparently free of maternal bias (Rhythmicity) failed to correlate with either questionnaire measure, lending support to the contention that the RITQ is measuring primarily maternal perceptions. There was little evidence in this study to support the contention that the RITQ does measure the infant's actual behavior.

However, there was ample evidence to support the validity of certain specific items on the RITQ, as specific items within scales matched very closely Diary items or clusters of items designed specifically to tap these contexts, even on scales with no overall correlation. Thus it is possible that the RITQ contains items that are irrelevant or nonrepresentative to the dimension measured, or that do not generate enough variance by virtue of their wording to be psychometrically useful.

Relationship Between RITQ and ICQ

Although the correlation between RITQ scores and corresponding ICQ factors was not a focus of this study, the finding that RITQ scales correlated at a significant level with corresponding ICQ factors lends support to the RITQ's convergent construct validity. The two measures do seem to be measuring similar constructs. This finding is consistent with previous studies finding moderate correlations between the RITQ and the ICQ (Goldsmith, Reiser-Danner & Briggs, 1991; Bates, 1979).

Observer Scores

The observer scores did not correlate at a significant level with either questionnaire or with the Diary, although correlations with the Diary did approach significance. This result was not expected, though the trend of Diary scores matching observer scores more closely was in the expected direction. The lack of significance in these results may be attributable to the very small sample of behavior that the observer scores reflect. It was expected that observer scores would correlate at a higher level with Diary scores than with questionnaire scores, under the assumption that the Diary and the observer are both measuring behavior as it occurs and are less influenced by maternal perceptions and biases than are questionnaires. This hypothesis was partially supported.

Conclusions

This study provided scant evidence to support the contention that the RITQ measures infant characteristics or objectively observed behavior, and not primarily maternal perceptions of infant temperament. It did provide evidence to indicate that mothers can be accurate observers and reporters of behavior when criteria are made specific, and when the infant's behavior is not dependent on maternal variables. It also provided further evidence of convergent validity in that the RITQ correlated at a moderate level with the ICQ, a questionnaire also designed to assess the dimensions involved in the measurement of difficult temperament. This finding is not inconsistent with the finding that the RITQ lacks construct validity in that it does not appear to be measuring characteristics of the child. It is likely that the correlations between the RITQ and the ICQ are a result of their shared variance in that they are both measuring maternal perceptions of infant temperament.

Limitations of the Study

Given the small and homogenous sample in this study, generalizations of these results must be limited to white infants in intact families with educated mothers. The sample was also limited in its variability of RITQ scores, in that there was a smaller percentage of infants rated as difficult in the sample than would be expected in the

population at large. Results may have been different in a sample with more variability in RITQ scores.

This study was also limited in its measure of observer-rated behavior. Observations were limited to two brief observations, and though interrater reliability was high, the sample of behavior was not sufficiently representative, in that only two ratings were done on each dimension measured in each observation. More detailed recording and rating of behavior would have provided a more representative and thus more sensitive measure of observed behavior.

Suggestions for Further Research

Given the clinical usefulness of the RITQ and its popularity in research, efforts to improve its psychometric properties are warranted. This study provided evidence to indicate that certain items on the RITQ were valid in that they correlated highly with mothers' Diary report of infant behavior; other items failed to correlate with any of the expected measures. Further analysis of individual items on the RITQ could shed more light on this finding, and perhaps identify why certain items are not as useful. Individual item analysis (in a larger sample than the present one) may reveal more information about the psychometric properties of items that do not provide much discriminatory power or that do not relate well to the scale in question. Items could be judged as to their clarity of meaning and relevance by a homogenous sample of mothers, or compared on an individual

basis with other measures. Problem items could then be either omitted or revised.

Further work in statistically delineating factors on the RITQ is also needed. The only study thus far with an adequate sample size to justify factor analysis was based on an Australian sample. Given that temperament ratings have been demonstrated to vary between cultures, factors derived from ratings from an American population sample may be different than those obtained from an Australian sample. Additionally, factor analytic methods which allow some degree of intercorrelation between factors (such as oblique methods) could be experimented with, as it has been suggested that temperamental traits are not necessarily independent of each other.

Additionally, as suggested by Bates and Bayles (1984), objective and subjective components of maternal ratings of temperament deserve further study. In-depth analysis of small samples of infants and mothers including several measures of maternal, family, and infant variables derived from various measurement methods may shed light on the relationship between maternal characteristics, maternal ratings of temperament, and objectively observed infant behavior. Until the relationship between these variables is clarified through empirical research, maternal ratings of infant temperament will continue to be contaminated with the vaguely specified factor of maternal bias.

APPENDIX

Informed Consent Form

This research project is a study of infant personality (or temperament) and its measurement. It is being carried out by Donna Kitch, a graduate student at the University of Florida, for her doctoral dissertation. She is being supervised by Dr. Yvonne Brackbill in the Psychology Department.

We will be interviewing you briefly about your baby and then asking you to fill out some questionnaires that ask you to rate and record your baby's behavior.

Any information you give us is strictly confidential; your infant will be identified by a number only. Results of the study may be published, but only in group form, so that no one family can be identified. Your participation, though greatly appreciated, is strictly voluntary.

Donna Kitch
Graduate Student
Psychology Dept.
University of Florida
(904) 392-9915

I have read the above.

Signed_____

INTERVIEW> remind confidentiality,(i.e. your answers will be completely anonymous and confidential.)

Family #_____

Interviewer_____

Date_____

Baby's first name_____ Baby's sex____ Birthdate_____

Who takes care of ____ when you can't? (hrs per week in care of:)

paid sitter at home_____ sitter's home_____

Father_____relative(s)_____(Who?)_____

Child care center_____

How is this arrangement working out?

_____

_____

Hours per week mother works outside of home_____

Hrs/wk father out_____

(father's participation:)

_____

_____

What is your baby like? How would you describe your baby's personality? (in your own words)

_____

_____

_____

What was _____ like as a newborn?

_____

_____

Do you have _____ on a regular feeding schedule or do you just feed him/her when he/she is hungry?

_____

_____

How about naps? Do you put her to sleep at regular times or when he/she is tired?

_____

_____

How do you feel about schedules for babies in general?

_____

_____

How has _____'s health been? Has h/she been sick very often? How many times in the last 2 months has s/he been to the Doctor (for illness)?

_____

_____

How much experience did you have with babies before having yours? (relatives, friends, child care experience)

_____

_____

Mother's age____ Education_____ Occupation_____

Father's age____ Education_____ Occupation_____

How long lived together _____ Date married_____

Anyone else living in the home? #_____relationship_____

How long living at present address_____Own or rent_____

House: sq. feet_____ LR____ DR____ BRs____ Baths____ Other_____

OBSERVER RATING OF INFANT BEHAVIOR - VISIT__1 or 2_(circle) Subj #_____

A.) INFANTS RESPONSE TO INTERVIEWER:
Baby's response to interviewer (after interviewer greets mother and then infant by saying "hi, _____" in a moderate voice.) What is infants **initial** response to the interviewer? Does infant make any move to **approach** or **withdraw** from the interviewer? (within first few minutes)

1= cries and fusses, or actively withdraws from interviewer
2= frightened; clings to mother
3= cautious, doesn't smile; hides face or avoids eye contact
4= no response, doesn't respond to interviewers presence
5= smiles slightly but is passive and does not initiate interaction
6= friendly, initiates interaction but doesn't touch interviewer
7= very friendly, actively intitiates interaction, touches interviewer

COMMENTS:_____

_____

Baby's response to interviewer after about 20 minutes have passed (but before toy is introduced):

1= still fussy and withdrawn, staying away from interviewer
2= still seems frightened and clinging to mother
3= cautious, doesn't smile; hiding face and avoiding eye contact
4= no response, doesn't respond to interviewers presence
5= smiling slightly but is passive and has not initiated interaction
6= friendly, has initiated interaction but has not touched interviewer
7= very friendly, has actively intitiated interaction or touched int.

COMMENTS:_____

_____

Emotional tone: rate second 15 minutes of interview, predominant mood: (before toy introduced)

1= laughing, excited movement of arms and legs
2= grinning, cooing
3= quietly content, slight smile
4= equally happy and fussy
5= more fussy than happy; complaining, somewhat irritable
6= fussy, not easily soothed, crying spells of short duration
7= very unhappy or upset; intense, loud crying

COMMENTS:_____

_____

_____

Subj #_____

B.) INFANTS RESPONSE TO NEW TOY: FIRST VISIT Toy used_____

Baby's response to new toy (after interviewer brings in toy and puts in front of baby) What is infants **initial** response to the toy? Does infant make any move to **approach** or **withdraw** from the toy?

1= cries and fusses, crawls away from toy
2= seems frightened or unhappy; moves away from toy
3= cautious, doesn't smile; doesn't touch toy
4= no response, doesn't seem to care about toy
5= smiles slightly, hestiantly moves to touch toy
6= smiles, quietly begins to play with toy
7= excited, laughs, plays with toy readily

COMMENTS:_____

_____

Note infants reaction to toy after 15 minutes:

1= still crying and fussing
2= frightened or unhappy; has not touched toy
3= cautious, hasn't smiled; hasn't touched toy
4= no response, doesn't seem to care about toy
5= smiles slightly, has hesitantly played with toy
6= smiling, quietly playing with toy
7= excited, laughing, playing with toy

COMMENTS:_____

_____

Emotional tone: rate overall mood while playing with toy:

1= laughing, excited movement of arms and legs
2= grinning, cooing
3= quietly content, slight smile
4= equally happy and fussy
5= more fussy than happy; complaining, somewhat irritable
6= fussy, not easily soothed, crying spells of short duration
7= very unhappy or upset; intense, loud crying

COMMMENTS:_____

_____

_____

(note what mother says about how typical baby's behavior today is)

Instructions for Introduction of Diary

After mother finishes questionnaires, is asked for feedback and thanked for filling them out, something similar to the following is said:

Since these questionnaires are somewhat limited in that they ask you to rate certain behaviors, we would like to get a more accurate picture of your baby's personality (temperament) by asking you to write down for several days what your baby actually does. We will be sending you a diary form that you can fill out for a few days. This way we can see how accurate these questionnaires are in the way that they ask you to describe your baby. You should receive this in the mail next week, and I will be calling you to answer any questions you might have about it.

EXPLANATION OF DIARY

The purpose of this diary is to give us an objective account of how your baby actually behaves at home every day. We think this method will give us more information than just observing your baby ourselves, since babies react differently to different people and places and can be quite unpredictable. We want to know what babies actually do, at home, on a daily basis.

We would like for you to keep this diary for a total of four days: two or three weekdays (Monday-Friday) and one or two weekend days (Saturday & Sunday). We realize that your schedule may not allow you to do this on consecutive days, so any four days within a 2 week period will be fine. Just be sure to write the date on each diary page.

There are two pages for each day in this diary. The first page is a chart, where you record the time of day when your baby eats, sleeps or gets changed (bowel movements only). The first column is to note the time, within a fifteen minute interval (for example 5:15 P.M.). The second column is to note your baby's mood at this time. (The scale for mood is listed on a separate sheet; you will be referring to it repeatedly, so we suggest you put it up on your refrigerator, or wherever you plan to keep the diary form). The third column is for any comments you may have.

The second page for each day is self-explanatory: each day you are asked to do something and note your baby's response. There are also a few questions to be answered at the end of the day, or late in the day. If your baby's response to something doesn't fit the choices, just describe what he/she did in your own words. Please write as many comments as you can on the diary, to help us understand what your baby is actually doing. Feel free to write on the sides, on the back, or on another sheet of paper.

If you have any questions at all, please call one of us. Your participation is very important to us, and we really appreciate your time and effort.

Donna Kitch
392-9915 (office)
or 338-1830 (home)

MOOD SCALE:

1 = laughing, excited movement of arms and legs

2 = grinning, cooing

3 = quietly content, slight smile

4 = neither happy nor upset

5 = more fussy than happy; complaining, somewhat irritable

6 = fussy, not easily soothed, crying spells of short duration

7 = very unhappy or upset; intense, loud crying

INSTRUCTIONS:

In the MOOD column on the first page of each day, record the number corresponding to the description that best fits your baby's mood during that time period. If one does not fit, for example if your baby was crying and then happy a few minutes later, put down both numbers, and make a note in the comments section.

DAY NUMBER ONE.
(day of week)_____ DATE:_____ FAMILY #:_____

1.) Did baby wake up last night? Circle NO YES #times_____

If YES, describe what happened: how long awake, whether fed, easily soothed, crying or playful, etc.

_____

_____

2.) TODAY, introduce your baby to a new vegetable, watch how your baby responds to it, and choose the description below that most accurately describes your baby's initial reaction to the new food.

vegetable chosen_____

__Hates it very much: spits it out, swats the spoon, cries or yells loudly.
__Hates it: makes a face, refuses more, turns head away.
__Doesn't particularly like it: swallows it but isn't pleased.
__Can't tell if they like it or not; no reaction.
__Seems to like it alright; accepts it but doesn't ask for more.
__Likes it: Smiles, is pleased, and readily accepts more.
__Loves it: Makes happy sounds, reacnes out or yells for more.

Comments_____

_____

3.) Did your baby meet any new or unfamiliar people today? NO YES
If so, what was baby's reaction? Circle one.

__ cried and fussed, or actively withdrew from stranger
__ frightened; clung to mother
__ was cautious, didn't smile; hid face or avoided eye contact
__ no response, didn't respond to stangers presence
__ smiled slightly but was passive and did not initiate interaction
__ friendly, initiated interaction but didn't touch stranger
__ very friendly, actively initiated interaction, touched stranger

4.) Did your baby do anything new or different today? (New food, new place, new toy, anything you can think of) Describe what happened and rate baby's first reaction, and eventual response (after used to it).

what happened?_____

First reaction:

loved it	liked it	neutral	didn't like it	hated it

Later on:

loved it	liked it	neutral	didn't like it	hated it

5.) Was baby's behavior typical today? (Describe anything you think may have affected baby's mood, i.e. teething, illness, your own mood, anything out of the ordinary going on at home) (write on back)

DAY NUMBER ONE
(day of week)_____ DATE:_____ FAMILY #:_____

	TIME:	MOOD #	COMMENTS
SLEEP: 1st awake time:			
1st nap asleep:			
awake:			
2nd nap asleep:			
awake:			
3rd nap asleep:			
awake:			

FEEDING:	TIME	MOOD	COMMENTS (include what eaten)
1st meal start:			
stop:			
2nd meal start:			
stop:			
3rd meal start:			
stop:			
4th meal start:			
stop:			
5th meal start:			
stop:			

DIAPERING: (record bowel movements only)

	TIME	MOOD	COMMENTS
first:			
second:			
third:			

BATHTIME:	TIME	MOOD	COMMENTS

DAY NUMBER TWO
(day of week)_____ DATE:_____ FAMILY #:_____

1.) Did baby wake up last night? Circle NO YES #times_____

If YES, describe what happened: how long awake, whether fed, easily soothed, crying or playful, etc.

_____

_____

TODAY, introduce your baby to a new napping place (sofa, play pen, different bed, anywhere s/he hasn't slept before), watch how your baby responds to it, and choose the description below that most accurately describes your baby's initial reaction to the new sleeping arrangement.

__Hates it very much: cries or yells loudly, refuses to stay.
__Hates it: fusses, makes initial attempts to get away
__Doesn't particularly like it: fusses a little, isn't pleased but stays.
__Can't tell if they like it or not; no reaction.
__Seems to like it alright; accepts it and lies down readily.
__Likes it: Smiles, is pleased, and goes to sleep readily.
__Loves it: Makes happy sounds, coos,and lies down smiling.

Comments_____

_____

3.) Did your baby meet any new or unfamiliar people today? NO YES
If so, what was baby's reaction? Circle one.

__ cried and fussed, or actively withdrew from stranger
__ frightened; clung to mother
__ was cautious, didn't smile; hid face or avoided eye contact
__ no response, didn't respond to stangers presence
__ smiled slightly but was passive and did not initiate interaction
__ friendly, initiated interaction but didn't touch stranger
__ very friendly, actively initiated interaction, touched stranger

4.) Did your baby do anything new or different today? (New food, new place, new toy, anything you can think of) Describe what happened and rate baby's first reaction, and eventual response (after used to it).

what happened?_____

First reaction:

loved it	liked it	neutral	didn't like it	hated it

Later on:

loved it	liked it	neutral	didn't like it	hated it

5.) Was baby's behavior typical today? (Describe anything you think may have affected baby's mood, i.e. teething, illness, your own mood, anything out of the ordinary going on at home) (write on back)

DAY NUMBER TWO
(day of week)_____ DATE:_____ FAMILY #:_____

	TIME:	MOOD #	COMMENTS
SLEEP: 1st awake time:			
1st nap asleep:			
awake:			
2nd nap asleep:			
awake:			
3rd nap asleep:			
awake:			

FEEDING:	TIME	MOOD	COMMENTS (include what eaten)
1st meal start:			
stop:			
2nd meal start:			
stop:			
3rd meal start:			
stop:			
4th meal start:			
stop:			
5th meal start:			
stop:			

DIAPERING: (record bowel movements only)

	TIME	MOOD	COMMENTS
first:			
second:			
third:			

BATHTIME:	TIME	MOOD	COMMENTS

DAY NUMBER THREE
(day of week)_____ DATE:_____ FAMILY #:_____

1.) Did baby wake up last night? Circle NO YES #times_____

If YES, describe what happened: how long awake, whether fed, easily soothed, crying or playful, etc.

_____

_____

TODAY, try the new vegetable again and record baby's reaction.

__Hates it very much: spits it out, swats the spoon, cries or yells loudly.
__Hates it: makes a face, refuses more, turns head away.
__Doesn't particularly like it: swallows it but isn't pleased.
__Can't tell if they like it or not: no reaction.
__Seems to like it alright; accepts it but doesn't ask for more.
__Likes it: Smiles, is pleased, and readily accepts more.
__Loves it: Makes happy sounds, reaches out or yells for more.

Comments:_____

_____

_____

3.) Did your baby meet any new or unfamiliar people today? NO YES
If so, what was baby's reaction? Circle one.

__ cried and fussed, or actively withdrew from stranger
__ frightened; clung to mother
__ was cautious, didn't smile; hid face or avoided eye contact
__ no response, didn't respond to stangers presence
__ smiled slightly but was passive and did not initiate interaction
__ friendly, initiated interaction but didn't touch stranger
__ very friendly, actively initiated interaction, touched stranger

4.) Did your baby do anything new or different today? (New food, new place, new toy, anything you can think of) Describe what happened and rate baby's first reaction, and eventual response (after used to it).

what happened?_____

First reaction:

loved it	liked it	neutral	didn't like it	hated it

Later on:

loved it	liked it	neutral	didn't like it	hated it

5.) Was baby's behavior typical today? (Describe anything you think may have affected baby's mood, i.e. teething, illness, your own mood, anything out of the ordinary going on at home) (write on back)

DAY NUMBER THREE
(day of week)_____ DATE:_____ FAMILY #:_____

TIME: MOOD # COMMENTS

SLEEP: 1st awake time:_____|_____|_____
1st nap asleep:_____|_____|_____
awake: _____|_____|_____
2nd nap asleep:_____|_____|_____
awake:_____|_____|_____
3rd nap asleep:_____|_____|_____
awake:_____|_____|_____

FEEDING: TIME MOOD COMMENTS (include what eaten)
1st meal start: _____|_____|_____
stop: _____|_____|_____
2nd meal start:_____|_____|_____
stop: _____|_____|_____
3rd meal start:_____|_____|_____
stop:_____|_____|_____
4th meal start:_____|_____|_____
stop:_____|_____|_____
5th meal start:_____|_____|_____
stop:_____|_____|_____

DIAPERING: (record bowel movements only)
TIME MOOD COMMENTS
first:_____|_____|_____
second:_____|_____|_____
third: _____|_____|_____

BATHTIME: TIME MOOD COMMENTS
_____|_____|_____

DAY NUMBER FOUR
(day of week)_____ DATE:_____ FAMILY #:_____

Did baby wake up last night? Circle NO YES #times_____

If YES, describe what happened: how long awake, whether fed, easily soothed, crying or playful, etc.

_____

_____

TODAY, try the new sleeping place again and record baby's reaction.

__Hates it very much: cries or yells loudly, refuses to stay.
__Hates it: fusses, makes initial attempts to get away
__Doesn't particularly like it: fusses a little, isn't pleased but stays.
__Can't tell if they like it or not; no reaction.
__Seems to like it alright; accepts it and lies down readily.
__Likes it: Smiles, is pleased, and goes to sleep readily.
__Loves it: Makes happy sounds, coos,and lies down smiling.

Comments:_____

_____

_____

3.) Did your baby meet any new or unfamiliar people today? NO YES
If so, what was baby's reaction? Circle one.

__ cried and fussed, or actively withdrew from stranger
__ frightened; clung to mother
__ was cautious, didn't smile; hid face or avoided eye contact
__ no response, didn't respond to stangers presence
__ smiled slightly but was passive and did not initiate interaction
__ friendly, initiated interaction but didn't touch stranger
__ very friendly, actively initiated interaction, touched stranger

4.) Did your baby do anything new or different today? (New food, new place, new toy, anything you can think of) Describe what happened and rate baby's first reaction, and eventual response (after used to it).

what happened?_____

First reaction:

loved it	liked it	neutral	didn't like it	hated it

Later on:

loved it	liked it	neutral	didn't like it	hated it

5.) Was baby's behavior typical today? (Describe anything you think may have affected baby's mood, i.e. teething, illness, your own mood, anything out of the ordinary going on at home) (write on back)

DAY NUMBER FOUR
(day of week)_____ DATE:_____ FAMILY #:_____

	TIME:	MOOD #	COMMENTS
SLEEP: 1st awake time:			
1st nap asleep:			
awake:			
2nd nap asleep:			
awake:			
3rd nap asleep:			
awake:			

FEEDING:	TIME	MOOD	COMMENTS (include what eaten)
1st meal start:			
stop:			
2nd meal start:			
stop:			
3rd meal start:			
stop:			
4th meal start:			
stop:			
5th meal start:			
stop:			

DIAPERING: (record bowel movements only)

	TIME	MOOD	COMMENTS
first:			
second:			
third:			

BATHTIME:	TIME	MOOD	COMMENTS

REFERENCE LIST

AERA, APA, & NCME (1985). Standards for educational and psychological testing. Washington, D.C.: American Psychological Association.

Affleck, G., Allen, D., McGrade, J. & McQueeney, M. (1983). Maternal and child characteristics associated with mothers' perceptions of their high risk/developmentally delayed infants. The Journal of Genetic Psychology, 142, 171-180.

Ainsworth, M.D.S., & Wittig, B.A. (1969). Attachment and exploratory behavior of one year olds in a Strange Situation. In B.M. Foss, (Ed.), Determinants of infant behavior, Vol. 4. London: Methuen.

Barnett, D.W. & MacMann, G.M. (1990). Personality assessment: Critical issues for research and practice. In C.R. Reynolds and R.W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Personality, behavior, and context. London: Guilford Press.

Bates, J.E. (1980). The concept of difficult temperament. Merrill-Palmer Quarterly, 26, 89-97.

Bates, J.E. (1986). The measurement of temperament. In R. Plomin & J. Dunn (Eds.), The study of temperament: Changes, continuities, and challenges. Hillsdale, N.J.: Erlbaum.

Bates, J.E. (1987). Temperament in infancy. In J.D. Osofsky (Ed.), Handbook of infant development (2nd ed.). New York: Wiley.

Bates, J.E. (1989). Concepts and measures of temperament. In G.A. Kohnstamm, J.E. Bates, & M.K. Rothbart (Eds.), Temperament in childhood. Chichester: Wiley.

Bates, J.E., & Bayles, K. (1984). Objective and subjective components in mother's perceptions of their children from age six months to three years. Merrill-Palmer Quarterly, 30, 111-130.

Bates, J.E. & Bayles, K. (1988). The role of attachment in the development of behavior problems. In J. Belsky & T. Nezworski, (Eds.), Clinical implications of attachment. New York: Erlbaum.

Bates, J.E., Freeland, C.B. & Lounsbury, M.L. (1979). Measurement of infant difficultness. Child Development, 50, 794-803.

Bates, J.E., Olson, S.L, Pettit, G.S. & Bayles, K. (1982). Dimensions of individuality in the mother-infant relationship at six months of age. Child Development, 53, 446-461.

Bornstein, M.H., Gaughran, J.M. & Homel, P. (1986). Infant temperament: theory, tradition, critique, and new assessments. In C.E. Izard & P.B. Read (Eds.), Measuring emotions in infants and children. Cambridge: Cambridge University Press.

Buss, A.H. & Plomin, R. (1984). Temperament: Early developing personality traits. Hillsdale, NJ: Erlbaum.

Brackbill, Y., White, M., Wilson, M., & Kitch, D. (1990). Family dynamics predict infant disposition. Infant Mental Health Journal, 11, 113-126.

Cameron, J. (1978). Parental treatment, children's temperament and the risk of childhood behavior problems. American Journal of Orthopsychiatry, 48, 140-147.

Cameron, J.R., Rice, D., Hansen, R. & Rosen, D. (1992, October). How infant temperament "works": A causal model analysis of the interrelationships between infant temperament scales. Paper presented at the Ninth Occasional Temperament Conference, Bloomington, IN.

Campos, J., Barrett, K., Lamb, M., Goldsmith, H., & Stenberg, C. (1983). Socioemotional development. In. P.H. Mussen (Ed.), Handbook of child psychology, (Vol. 2, 4th ed.). New York: Wiley.

Carey, W.B. (1982). Validity of parental assessments of development and behavior. American Journal of the Diseases of Children, 136, 97-99.

Carey, W.B. (1985a). Temperament and increased weight gain in infants. Developmental and Behavioral Pediatrics, 6, 128-131.

Carey, W.B. (1985b). Clinical use of temperament data in pediatrics. Journal of Developmental and Behavioral Pediatrics, 6, 137-142.

Carey, W.B. (1989). Practical applications in pediatrics. In G.A. Kohnstamm, J.E. Bates & M.K. Rothbart (Eds.), Temperament in childhood. Chichester: Wiley.

Carey, W.B. & McDevitt, S.C. (1978). Revision of the Infant Temperament Questionnaire. Pediatrics, 61, 735-739.

Chen, C., Yu, M., Wang, C. and Tong, S. (1990). Genetic variance and heritability of temperament among Chinese twin infants. Acta Geneticae Medicae et Gemellologiae Twin Research, 39(4), 485-490.

Cutrona, B. & Troutman, B. (1986). Social support, infant temperament and parenting self-efficacy. Child Development, 57, 1507-1518.

Cyphers, L., Phillips, K., Fulker, D. & Mrazek, D. (1990). Twin temperament during the transition from infancy to early childhood. Journal of the American Academy of Child and Adolescent Psychiatry, 29(3), 392-397.

DeVries, M. (1984). Temperament and infant mortality among the Masai of East Africa. American Journal of Psychiatry, 141, 1189-1194.

DiBlasio, C., Bond, L., Wasserman, R., & Creasey, G. (1988, April). Infant temperament and behavior problems at six to seven years. Paper presented at the International Conference of Infant Studies. Washington, D.C..

Frodi, A., Bridges, L., & Shonk, S. (1989). Maternal correlates of infant temperament ratings and of infant-mother attachment: A longitudinal study. Infant Mental Health Journal, 10(4), 273-289.

Goldsmith, H.H., Buss, A.H., Plomin, R., Rothbart, M.K., Thomas, A., Chess, S., Hinde, R.A., & McCall, R.R. (1987). Roundtable: What is temperament? Four approaches. Child Development, 58, 505-529.

Goldsmith, H.H. & Campos, J.J. (1982). Toward a theory of infant temperament. In R.N. Emde & R.J. Harmon (Eds.), The development of attachment and affiliative systems. New York: Plenum.

Goldsmith, H.H., Elliot, T.K., & Jaco, K.L. (1986). Construction and initial validation of a new temperament questionnaire. Infant Behavior and Development, 9, 144.

Goldsmith, H.H. & Reiser-Danner, L.A. (1990). Assessing early temperament. In C.R. Reynolds & R.W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Personality, behavior, and context. London: Guilford Press.

Goldsmith, H.H., Rieser-Danner, L.A., & Briggs, S. (1991). Evaluating convergent and discriminant validity of temperament questionnaires for preschoolers, toddlers, and infants. Developmental Psychology, 27(4), 566-579.

Gunnar, M., Larson, M., Hertsgaard, L., Harris, M., & Brodersen, L. (1992). The stressfulness of separation among nine-month old infants: Effects of social context variables and infant temperament. Child Development, 63, 290-303.

Hara, H., Mitsuishi, C., & Yamaguchi, K. (1990). Temperament of very-low-birthweight infants. Journal of Mental Health, 36, 79-84.

Healy, B. (1989). Autonomic nervous system correlates of temperament. Infant Behavior and Development, 12(3), 289-304.

Hegvik, R.L., McDevitt, S.C. & Carey, W.B. (1980). The Middle Childhood Temperament Questionnaire. Journal of Developmental and Behavioral Pediatrics, 3, 197-200.

Houldin, A. (1987). Infant temperament and the quality of the childrearing environment. Maternal Child Nursing Journal, 16(2), 131-143.

Kemp, V. (1987). Mothers' perceptions of childrens' temperament and mother-child attachment. Scholarly Inquiry for Nursing Practice, 1(1), 51-68.

Kyrios, M., Prior, M., & Oberklaid, F. (1989). Cross-cultural studies of temperament: Temperament in Greek infants. International Journal of Psychology, 24(5), 585-603.

Koniak-Griffin, D. & Rummell, M. (1988). Temperament in infancy: Stability, change, and correlates. Maternal Child Nursing Journal, 17(1), 25-40.

Lounsbury, M. & Bates, J. (1982). The cries of infants of differing levels of perceived temperamental difficultness: Acoustic properties and effects on listeners. Child Development, 53, 677-686.

Matheny, A., Wilson, R., & Nuss, S. (1984). Toddler temperament: Stability across settings and over ages. Child Development, 55, 1200-1211.

McDevitt, S.C., & Carey, W.B. (1975). The measurement of temperament in 3-7 year old children. Journal of Child Psychology and Psychiatry, 19, 245-253.

McDevitt, S.C., & Carey, W.B. (1981). Stability of ratings versus perceptions of temperament from early infancy to 1-3 years. American Journal of Orthopsychiatry, 51, 342-345.

McNeil, T., & Persson-Blennow, I. (1988). Stability of temperament characteristics in childhood. American Journal of Orthopsychiatry, 58(4), 622-626.

Medoff-Cooper, B. (1986). Temperament in very low birth weight infants. Nursing Research, 35(3), 139-143.

Nunnally, J. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Nyman, G. (1988). Infant temperament, childhood accidents, and hospitalization. Annual Progress in Child Psychiatry and Child Development, 415-429.

Oberklaid, F., Prior, M., Nolan, T., & Smith, P. (1985). Temperament in infants born prematurely. Journal of Developmental and Behavioral Pediatrics, 6(2), 57-61.

Peters-Martin, P. & Wachs, T.D. (1984). A longitudinal study of temperament with its correlates in the first 12 months. Infant Behavior and Development, 7, 285-298.

Power, T.G., Gershenhorn, S., & Stafford, D. (1990). Maternal perceptions of infant difficultness: the influence of maternal attitudes and attributions. Infant Behavior and Development, 13, 421-437.

Priel, B., Henik, A., Dekel, A., & Tal, A. (1990). Perceived temperamental characteristics and regulation of physiological stress: A study of wheezy babies. Journal of Pediatric Psychology, 15(2), 197-209.

Prior, M. (1992). Childhood temperament. Journal of Child Psychology and Psychiatry, 33, 249-279.

Prior, M., Kyrios, M., & Oberklaid, F. (1986). Temperament in Australian, Chinese, and Greek infants. Journal of Cross-Cultural Psychology, 17, 455-474.

Prior, M., Oberklaid, F., & Sanson, A. (1987). The Short Temperament Scale for Infants. Bundoora, Australia: The Australian Temperament Project.

Prior, M., Sanson, A. & Oberklaid, F. (1989). The Australian temperament project. In G.A. Kohnstamm, J.E. Bates & M.K. Rothbart (Eds.), Temperament in childhood. Chichester: Wiley.

Rennen-Allhoff, B., & Reinhard, H. (1988). An investigation of German versions of the temperament questionnaires developed by Carey and co-workers. Zeitschrift fur Kinder und Jugendpsychiatrie, 16(2), 61-66.

Rice, M.S. & Gaines, S.K. (1992). Measurement of child temperament: Implications for researchers, clinicians, and caregivers. Children's Health Care, 21(3), 177-183.

Richards, J. & Cameron, D. (1989). Infant heart-rate variability and behavioral developmental status. Infant Behavior and Development, 12(1), 45-58.

Roth, K., Eisenberg, N. & Seu, E. (1984). The relation of preterms and full terms infants' temperament to test-taking behaviors and developmental status. Infant Behavior and Development, 7, 495-505.

Rothbart, M.K. (1982). The concept of difficult temperament: A response to Thomas, Chess, and Korn. Merrill-Palmer Quarterly, 28, 35-40.

Rothbart, M.K. (1986). A psychobiological approach to the study of infant temperament. In G.A. Kohnstamm (Ed.), Temperament discussed. Lisse: Swets & Zeitlinger.

Rothbart, M.K. (1988). Temperament and the development of inhibited approach. Child Development, 59, 1241-1250.

Rothbart, M.K. & Goldsmith, H.H. (1985). Three approaches to the study of infant temperament. Developmental Review, 5, 237-260.

Sameroff, A.J., Seifer, R., & Elias, R. (1982). Sociocultural variability in infant temperament ratings. Child Development, 53, 164-173.

Sanson, A., Prior, M., & Oberklaid, F. (1985). Normative data on temperament in Australian infants. Australian Journal of Psychology, 37, 185-195.

Sanson, A., Prior, M., Garino, E., Oberklaid, F. & Sewell, J. (1987). The structure of infant temperament: Factor analysis of the Revised Infant Temperament Questionnaire. Infant Behavior and Development, 10, 97-104.

Shaefer, C. (1990). Night waking and temperament in early childhood. Psychological Reports, 67(1), 192-194.

Slabach, E.H., Morrow, J., & Wachs, T.D. (1991). Questionnaire measurement of infant and child temperament. In J. Strelau and A. Angleitner (Eds.), Explorations in temperament. London: Plenum Press.

Sprunger, L.W., Boyce, W.T., & Gaines, J.A. (1985). Family-infant congruence: Routines and rhythmicity in family adaptations to a young infant. Child Development, 56, 564-572.

Spungen, L. & Farran, A. (1986). Effect of intensive care unit exposure on temperament in low birth weight preterm infants. Journal of Developmental and Behavioral Pediatrics, 7(5), 288-292.

Strelau, J. & Angleitner, A. (1991). Temperament research: some divergencies and similarities. In J. Strelau and A. Angleitner (Eds.), Explorations in temperament. London: Plenum Press.

Thomas, A. & Chess, S. (1977). Temperament and development. New York: Brunner/Mazel.

Thomas, A., Chess, S., & Birch, H.G. (1968). Temperament and behavior disorders in children. New York: New York University Press.

Thomas, A., Chess, S., & Korn, S.J. (1982). The reality of difficult temperament. Merrill-Palmer Quarterly, 28, 21-24.

Tomasdottir, M., Wilson, M.E., White, M.A. & Agustsdottir, T. (1991). Family dynamics and infant temperament in urban Iceland. Scandinavian Journal of Caring Science, 5, 211-217.

Vaughn, B.E., Bradley, C.F., Joffe, L.S., Seifer, R. & Barglow, T. (1987). Maternal characteristics measured prenatally are predictive of ratings of temperamental "difficulty" on the Carey Infant Temperament Questionnaire. Developmental Psychology, 23, 152-161.

Vaughn, B., Lefever, G., Seifer, R., & Barglow, P. (1989). Attachment behavior, attachment security, and temperament during infancy. Child Development, 60(3), 728-737.

Vaughn, B.E., Taraldson, B.J., Crichton, L. & Egeland, B. (1981). The assessment of infant temperament: A critique of the Carey Infant Temperament Questionnaire, Infant Behavior and Development, 4, 1-17.

Watt, J. (1987). Temperament in small-for-dates and pre-term infants: A preliminary study. Child Psychiatry and Human Development, 17(3), 177-188.

Weissbluth, M. (1981). Sleep duration and infant temperament. Journal of Pediatrics, 99, 817-819.

Windle, M. (1988). Psychometric strategies of measures of temperament: A methodological critique. International Journal of Behavioral Development, 11, 171-201.

Zeanah, C., Keener, M., & Anders, T. (1986). Developing perceptions of temperament and their relation to mother and infant behavior. Journal of Child Psychology and Psychiatry, 27, 499-512.

Zeanah, C., Keener, M., Anders, T., & Vieira-Baker, C. (1987). Adolescent mothers' perceptions of their infants before and after birth. American Journal of Orthopsychiatry, 57(3), 351-360.

BIOGRAPHICAL SKETCH

Donna S. Kitch was born Donna S. Hussey on September 13, 1955 in Pensacola, Florida. She graduated from Weston High School in Weston, Connecticut in 1973, and received a Bachelor of Science degree in mental health from Georgia State University in 1979.

She completed her predoctoral internship at the Child and Adolescent Service Center in Canton, Ohio in 1992, and expects to receive her Ph.D. in counseling psychology from the University of Florida in August, 1993.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Franz Epting, Chair
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Yvonne Brackbill, CoChair
Graduate Research Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Margaret Wilson
Associate Professor of Nursing

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry Grater
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

S. Graham Kosch
Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the Department of Psychology in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1993

_____
Dean, Graduate School